MULTIVARIATE META-ANALYSIS METHODS FOR BIAS REDUCTION IN SYSTEMATIC REVIEWS

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Giacomo Frosi

ACKNOWLEDGMENTS

I would like to thank Professor Paula Williamson, Dr Jamie Kirkham and Professor Richard Riley for their assistance in the preparation of this thesis and for all their help, advice and support during my PhD. My gratitude also goes to the members of the department, for their advice and their assistance especially to Dr Steven Lane, Dr Susanna Dodd, Dr Laura Bonnett (PhD student mentor) for their assistance with specific topics and for their encouragement.

My personal thanks goes to my parents, Rosa and Alberto, for their support, encouragement and advice. I also would like to thank all my dear friends, with a special thanks to Margaret, Giovanna and Luca for their patience, encouragement and understanding.

DECLARATION

This is all my own work and I have used my own computer codes where applicable. In this thesis I refer to the ORBIT study. This study was a collaborative project conducted by my two supervisors Paula Williamson and Jamie Kirkham in collaboration with Professor Doug Altman, Dr Carrol Gamble, Professor Ann Jacoby, Dr Kerry Dwan, Dr Rebecca Smyth and Dr Susanna Dodd.

The paper resulting from this work:

Frosi G, Riley RD, Williamson PR, Kirkham JJ. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. Journal of Clinical Epidemiology, (2015).

ABSTRACT

Background: Missing treatment effect estimates for particular outcomes in a study have the potential to affect the conclusions in a meta-analysis (MA), especially if missingness is a result of outcome reporting bias (ORB). ORB comes from the results-based selection for publication of a subset of the original measured outcome variables and can result in overestimating treatment effects, resulting in bias. As well as missing treatment effect estimates at the study level, outcome data may also be missing within studies at the individual participant level. Multivariate meta-analysis (MVMA) of individual participant data (IPD) has the potential to overcome the impact of both these problems, by utilising the correlation between outcomes.

Methods: An assessment of ORB was carried out in a cohort of systematic reviews (SR) with a core set of outcomes investigating pharmacological treatments for rheumatoid arthritis (RA). Novel IPD-MVMA methods to borrow strength across correlated outcomes were applied and evaluated through simulation, with the aim to show and quantify how this approach can reduce bias and improve precision of MA results, compared to traditional univariate methods, when there is missing outcome data as a result of both ORB and missing participant data. ORB assessments and the MVMA methods were applied to examples in RA.

Results: Of the 167 assessable trials from 21 Cochrane RA reviews, 23% contained high suspicion of ORB in at least one of the core outcomes. Results from the simulations showed that the 'borrowing of strength' (BoS) in a multivariate model can reduce the magnitude of bias and increase precision in the pooled estimates. Results showed that when an ORB mechanism is introduced or there was missing IPD, MVMA tends to reduce the bias, increase the precision and improve the coverage when compared to a univariate analysis. In some instances, these benefits observed were also found in applying MVMA to the RA reviews.

Conclusions: MVMA is not the solution to all missing data related problems within review meta-analyses, but, informed by this thesis, it can unquestionably be seen as a route to address missing outcome data in SRs. The BoS, reduction in bias and increase in precision can all be seen as promising.

List of Figures:

List of Tables:

Appendix A

Appendix C

Appendix E

Legend of Abbreviations:

ACR:	American College of Rheumatology
AD:	Aggregate data
AD-MA:	Aggregate data meta-analysis
APR:	Acute phase reactant
BoS:	Borrowing of strength
BFMA:	Bivariate fixed-effects meta-analysis
CDSR:	Cochrane Database of Systematic Reviews
CMSG:	Cochrane Musculoskeletal Group
COMET:	Core Outcome Measures in Effectiveness Trials
CONSORT:	Consolidated Standards of Reporting Trials
COS:	Core outcome set
CRGs:	Cochrane Review Groups
DAS:	Disease Activity Score
DMARD:	Disease-modifying antirheumatic drug
EBM:	Evidence-based medicine
EULAR:	European League Against Rheumatism
FEMA:	Fixed-effects meta-analysis
FR:	Fully reported
HAQ:	Health Assessment Questionnaire
IPD:	Individual participant data
IPD-MA:	Individual participant data meta-analyses
IQR:	Inter Quartile Range
MA:	Meta-analysis
MAR:	Missing at random
MCAR:	Missing completely at random
MD:	Mean Difference
MFMA:	Multivariate fixed-effects meta-analysis
ML:	Maximum likelihood
MNAR:	Missing not at random
MRMA:	Multivariate random-effects meta-analysis
MVMA:	Multivariate meta-analysis

NICE:	National Institute for Health and Clinical Excellence
NSAIDs:	Non-steroidal anti-inflammatory drugs
OA:	Osteoarthritis
OMERACT:	Outcome Measures for Rheumatology Clinical Trials
ORB:	Outcome reporting bias
ORBIT:	Outcome Reporting Bias in Trials
Pat.Global:	Patient global
Phy.Global:	Physician global
PRISMA:	Preferred Reporting Items for SRs and Meta-Analysis
RA:	Rheumatoid arthritis
RCT:	Randomised controlled trials
RD:	Radiological damage
REMA:	Random-effects meta-analysis
REML:	Restricted maximum likelihood
SE:	Standard error
SJC:	Swollen joint count
SR:	Systematic reviews
TJC:	Tender joint count
UFMA:	Univariate fixed-effects meta-analysis
URMA:	Univariate random-effects meta-analysis
UVMA:	Univariate meta-analysis
w/s:	within-study

Chapter 1 – Introduction

1.1 Evidence-based medicine

It is well established that evidence-based medicine (EBM) is a hot topic for clinicians, public health practitioners, purchasers, planners, and the public [1]. EBM is defined as the conscientious, explicit and judicious use of current best evidence in making decisions about the care of patients [1].

It is essential at this point of this thesis to begin with an understanding of what we mean by EBM.. The aim of EBM is to help physicians find the information that will ensure they can provide optimum management for their patients [2]. In essence, EBM consists of five linked ideas: identifying the clinical issues and questions, obtaining the clinical evidence, making recommendations from the evidence (evidence synthesis), making an informed clinical decision for patients and evaluating the performance of the decisions made [2].

In other words, the practice of EBM involves converting the need for information into a question and identifying the best evidence in order to answer that question. It consists of integrating individual clinical expertise with the best available external clinical evidence from systematic research [2].

Therefore, in accordance with the definition and its targets, it can be said, as suggested by Sackett et al. [1], that EBM is a hot topic for clinicians, public health practitioners, purchasers, planners and the public.

1.2 Randomised controlled trials

A randomised controlled trial (RCT) could be defined as a planned experiment involving the random assignment of participants to interventions [3]. For example, in a two-arm parallel group RCT of rheumatoid arthritis (RA), participants allocated to the treatment group may receive a biologic treatment (intervention) and those allocated to the alternative group may receive a standard disease-modifying anti-rheumatic drug (control). If an RCT is designed and conducted correctly, randomisation reduces bias by allowing the investigators to control for factors, both known and unknown, that would otherwise confound the results [3].

The strength of evidence for decisions about the effects of an intervention can be thought of as a pyramid (Figure 1.1 http://sladen.hfhs.org/library/staff/ebm-resource-pyramid.htm), with randomised trials and systematic reviews (SRs) of these providing the strongest evidence [1]. EBM incorporates the best available research evidence, and RCTs are seen as the gold standard of study designs to evaluate the effectiveness of a treatment in medical research in humans.

Figure 1.1 The Evidence-Based Medicine Pyramid of Evidence

Nevertheless, in some situations it may not be possible or ethical to conduct an RCT to answer a specific research question. For example, non-randomised observational studies

may be more appropriate to investigate serious, rare and long term harmful or unintended effects of treatment [3].

A report of an RCT should enable readers to understand the conduct of the trial and to assess the validity of its results. The Consolidated Standards of Reporting Trials (CONSORT) Statement is intended to improve the reporting of a randomised controlled trial [4]. It comprises a checklist of 22 essential items to be reported and a flow diagram.

1.3 Systematic reviews

When conducting a trial, similar studies may be considered alongside the trial but they are often analysed and reported in isolation [5]. In this section, we introduce the systematic review (SR) as a means to appraise and synthesise the research evidence from the individual studies that aim to answer the same question. Focus is given to the Cochrane Musculoskeletal Group (CMSG) SR for RA [6], which is considered in more detail as the motivating example of this thesis.

1.3.1 Need for systematic reviews in scientific research

Healthcare providers, consumers, researchers, and policy makers are inundated with unmanageable amounts of information, including evidence from healthcare research [1]. It is unlikely that all will have the time, skills and resources to find, appraise and interpret this evidence and to incorporate it into healthcare decisions. Cochrane reviews respond to this challenge by identifying, appraising and synthesising research-based evidence and presenting it in an accessible format [7].

A SR attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimising bias, thus providing more reliable findings from which conclusions can be drawn and decisions made [7]. The key characteristics of a SR are:

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;

- an explicit, reproducible methodology;

- a systematic search that attempts to identify all studies that would meet the eligibility criteria;

- an assessment of the validity of the findings of the included studies, for example, through the assessment of risk of bias; and

- a systematic presentation, and synthesis, of the characteristics and findings of the included studies.

Many SRs contain meta-analyses. Meta-analysis (MA) is the use of statistical methods to summarise the results of independent studies [7]. By combining information from all relevant studies, meta-analyses can provide more precise estimates of the effects of healthcare than those derived from the individual studies included within a review. They also facilitate investigation of the consistency of evidence across studies and the exploration of differences across studies.

1.3.2 Cochrane Collaboration and systematic reviews

The Cochrane Collaboration (http://www.cochrane.org) is an international organisation whose primary aim is to help people make well-informed decisions about healthcare by preparing, maintaining and promoting the accessibility of SRs of the evidence that underpins them. By providing a reliable synthesis of the available evidence on a given topic, SRs adhere to the principle that science is cumulative and facilitate decisions considering all the evidence on the effect of an intervention. Since it was founded in 1993, The Cochrane Collaboration has grown to include over 15,000 contributors from more than 100 countries, making it easily the largest organisation involved in this kind of work [7].

The work of The Cochrane Collaboration orbits around 53 Cochrane Review Groups (CRGs), responsible for preparing and maintaining reviews within specific areas of healthcare. Cochrane reviews are published online in the Cochrane Database of Systematic Reviews (CDSR), which is a core component of The Cochrane Library. The CDSR is published 12 times a year, each time with new reviews and updates of existing reviews. The Library currently contains (March 2011) more than 4500 Cochrane reviews and 2000 protocols for reviews in progress.

1.3.3 Cochrane Reviews of Rheumatoid Arthritis - Musculoskeletal Group

The CMSG represents one of the 53 CRGs. The targets of the CMSG include: gout, lupus erythematosus, osteoarthritis (OA), osteoporosis, paediatric rheumatology, RA, soft tissue rheumatism, spondyloarthropathy, systemic sclerosis and vasculitis.

As of May 2014, the CMSG had produced 171 reviews and 105 protocols. They had registered 15 titles, and contributed 1023 studies to the CENTRAL register (http://musculoskeletal.cochrane.org/welcome) edition of the CDSR.

The focus of this thesis is on RA, for which there were 59 published reviews as of Issue 9, 2012 (http://onlinelibrary.wiley.com/book/10.1002/14651858/topics).

1.4 Meta-analysis

It is now widely recognised that MA is a vital part of a SR in order to analyse, interpret and communicate a collection of clinical trials. For the last few years, MA has been of increasing interest in medicine. One of the principal reasons for conducting a MA within a SR is that by combining information from all relevant studies it increases the power and precision of estimates of the treatment effect [8].

The aim of this section is to explain what MA is and introduce the statistical models used. At this stage of the thesis we will provide the univariate model for MA where it is assumed that all outcomes (endpoints) are analysed separately.

In general, MA is a two-stage process [9]:

- Obtain the treatment effect estimate and variability (e.g. standard error (SE)) in each study.

- Combine the treatment effect estimates from each study using a weighted average to obtain a pooled estimate of treatment effect and appropriate confidence interval.

Assume that we have *n* studies (*i* =1 to *n*). The general formula for a pooled estimate is [10]:

$$\hat{y}_i = \frac{\sum_{i=1}^{n} w_i \mu_i}{\sum_{i=1}^{n} w_i} \qquad (1.1) \ .$$

Where $\hat{y}_i$ is the pooled treatment effect estimate, the weight is denoted as w_i and the specific study effect estimate as μ_i .

The studies in the MA are weighted so that studies that contain the most information and have estimated the treatment effect more precisely (narrow confidence intervals) are given the most weight. The weight is generally consistent with sample size such that larger studies have a larger weight and therefore a smaller confidence interval width [11]. For binary outcomes the confidence interval width also depends on the frequency of the event. Therefore, a smaller trial may have a larger weight if the frequency of events is high [11].

The weight given to a study is calculated as the reciprocal of the variance of the effect size,

$$w_i = \frac{1}{\sigma_i^2} \qquad (1.2).$$

In formula the variance is denoted as σ_i^2. Therefore, studies with more information have small variances and are given a larger weight [10].

1.4.1 Definition of heterogeneity and importance to assess it in systematic reviews

The studies that are brought together in a SR could differ for various reasons leading to variability among the studies. All the types of variability among studies in a SR can be defined as heterogeneity [9]. It is possible to differentiate between different types of heterogeneity. The first type of heterogeneity is due to the variability in the participants, interventions and outcomes studied. This is defined as *clinical diversity* [9]. A second type of heterogeneity is due to the variability in study design and risk of bias, and is defined as *methodological diversity* [9]. It is essential to assess these two types of heterogeneity, as only if studies are assessed to be clinical and methodological homogeneous is it possible to combine these studies in a MA and explore statistical heterogeneity [9].

1.4.2 Statistical heterogeneity

Statistical heterogeneity is one of the most troublesome aspects of many SRs. The interpretative problems depend on how substantial the heterogeneity is, since this determines the extent to which it might influence the conclusions of the MA. It is therefore important to be able to quantify the extent of heterogeneity among a collection of studies [9]. An obvious means of achieving this is by estimating the between-study variance of the parameters of interest [11]. This is performed as part of random-effects meta-analysis (REMA) [11]. There are two approaches for detecting heterogeneity: graphical and statistical.

The graphical approach is called a '*Forest plot*'; that is, a graphical display of results from individual studies on a common scale [12]. In this approach the effect estimate is usually represented by a square and the size of this square is representative of the weight the study has in the MA. Associated with each estimate there will be also a graphical visualisation of the confidence intervals calculated. Clearly, a less informative trial will have a wider confidence interval, meaning that the estimate of the treatment effect is less precise. This plot allows a visual examination of the degree of heterogeneity between studies. The more the confidence intervals overlap, the less heterogeneity there is.

Considering the statistical approach, there are two different methods. The first method is the heterogeneity statistic called Cochran's Q given by [13]:

$$Q = \sum w_i (\mu_i - \hat{\mu})^2 \qquad (1.3)\,.$$

In equation 1.3, $\hat{\mu}$ refers to the pooled true treatment effect estimate, while as described in the previous paragraph w_i denotes the weight, and μ_i denotes the specific study effect estimate.

P-values are obtained by comparing the statistic Q with a χ^2 distribution with n−1 degrees of freedom, where *n* is the number of studies. This test however is known to have low power to detect heterogeneity and it is suggested to use a value of 0.1 as a cut-off for significance. Indeed, low p-values (p<0.1) indicate statistical heterogeneity [13]. Conversely, Q has too much power as a test of heterogeneity if the number of studies is large [13].

An alternative method to calculate the effect of heterogeneity, providing a measure of the degree of inconsistency in the studies' results, is the I^2. As suggested by Higgins et al. [9], the I^2 statistic evaluates the percentage of total variation across the study that is not attributable to chance alone. The formula of Higgins and Thompson [11] to calculate I^2 will be explained and reported in details in Chapter 5 in section 5.3.6.1.

1.4.3 Quality of reporting in randomised controlled trials

One of the issues when conducting a MA is the quality of reporting of RCTs. In 1996 a group of clinical epidemiologists, clinicians, statisticians, editors and researchers met and set up the Preferred Reporting Items for SRs and Meta-Analysis (PRISMA) group [14], with the aim of improving the quality of MA reporting. Over the years, several medical journals have accepted the PRISMA statement to improve the reporting of SRs that they publish.

1.4.4 Fixed-effects meta-analysis

In fixed-effects meta-analysis (FEMA) it is assumed that all studies are estimating the same true effect. It is also assumed that the variability between studies results is due solely to the sample of people within each study. Another important assumption is that precision depends mainly on study size [15].

For each study *i,* a set of *m* treatment effects $(y_{i1} \quad y_{i2} \quad .... \quad y_{im})$ and their within-study (w/s) variances $(s_{i1}^2 \quad s_{i2}^2 \quad .... \quad s_{im}^2)$ are required.

In the univariate fixed-effects meta-analysis (UFMA) each outcome is analysed separately, and assumes that the obtained estimates of the treatment effect from the i^{th} study are normally distributed about a common fixed true effect ($\mu_{i1} \quad \mu_{i2} \quad .... \quad \mu_{im}$ respectively for outcomes 1, 2…, m) and variances, which are assumed to be known; therefore, for the i^{th} study and, the general UFMA could be written as follows [16]:

$$\begin{aligned} y_{i1} &\sim N\left(\mu_1, s_{i1}^2\right) \\ y_{i2} &\sim N\left(\mu_2, s_{i2}^2\right) \\ y_{im} &\sim N\left(\mu_m, s_{im}^2\right) \end{aligned} \qquad (1.5) .$$

In this model, we are interested in the estimate of the common fixed true effect. The model parameters are often estimated using the method of maximum likelihood (ML).

1.4.5 Random-effects meta-analysis

For REMA the model assumes that the treatment effects for the individual studies are assumed to vary around a global treatment effect.

In each study, the overall treatment effects $(y_{i1} \quad y_{i2} \quad .... \quad y_{im})$ are assumed to be an estimate of a true value $(\mu_{i1} \quad \mu_{i2} \quad .... \quad \mu_{im})$; and in a hierarchical structure each true value is assumed to follow a Normal distribution.

The general univariate random-effect meta-analysis (URMA) could be written as follows [17]:

$$\begin{cases} y_{i1} \sim N\left(\mu_1, s_{i1}^2\right) & \mu_{i1} \sim N\left(\beta_1, \tau_1^2\right) \\ y_{i2} \sim N\left(\mu_2, s_{i2}^2\right) & \mu_{i2} \sim N\left(\beta_2, \tau_2^2\right) \\ y_{im} \sim N\left(\mu_m, s_{im}^2\right) & \mu_{im} \sim N\left(\beta_m, \tau_m^2\right) \end{cases} \qquad (1.6)$$

In this model, the objective is to obtain the mean effect $(\beta_1 \quad \beta_2 \quad .... \quad \beta_m)$ and the between-studies $(\tau_1^2 \quad \tau_2^2 \quad .... \quad \tau_m^2)$ variances and possibly the true values-specific effects $(\mu_{i1} \quad \mu_{i2} \quad .... \quad \mu_{im})$ [17].

The model parameters are usually estimated using the method of restricted maximum likelihood (REML).

1.5 Individual participant data meta-analysis

The purpose of this section is to provide a general introduction of the concept of individual participant data meta-analyses (IPD-MA) approaches and their definitions. A more detailed

description will be given in the methodology chapters which will present the models and their formulas with reference to the simulation work conducted in this thesis.

The IPD-MA models are widely considered the gold standard of MA and involve the re-analysis of all the individual participant data (IPD) from the included studies [18]. Most often meta-analyses are aggregate but there are advantages to using IPD [18]. However, availability of IPD and lack of resources often mean that aggregate meta-analyses are more common [18]. In this thesis, both aggregate and IPD meta-analyses approaches are explored.

When IPD are available from each of the studies, the MA can proceed in either a one-step or a two-step framework. The *one-stage* approach simultaneously models the IPD from all of the studies. The *two-stage* approach first fits a model to the IPD from each study separately, and then the study parameter estimates are combined in a MA [19].

1.5.1 One-stage individual participant data meta-analysis

In the one-stage approach, the IPD from all studies are modelled simultaneously while accounting for the clustering of participants within studies.

A one-stage model [18] can be used to estimate intervention effects while stratifying or otherwise accounting for differences between trials [20]. This is typically a regression model such as linear, logistic, or Cox regression, with either a separate term for each trial or one that varies across trials via a random effect.

1.5.2 Two-stage individual participant data meta-analysis

In this section a general description of two-stage IPD-MA will be provided. Further details with formula and an application to the simulation that was conducted in this thesis work will be provided in Chapter 3 and Chapter 4.

- In the first stage of the two-step IPD-MA approach, each study is analysed separately.

- The second stage requires a univariate MA framework or a multivariate MA approach.

Two-stage IPD-MA are clearly more laborious, but in the second stage the models allow the use of traditional, well-known MA techniques such as those used by the Cochrane Collaboration (for example, inverse variance FE or RE approach, or the Mantel-Haenszel method).

One-stage IPD-MA approach and two-stage IPD–MA approach could lead to different results due to some assumptions that will be described in detail in Chapter 3 of this thesis.

In this work of thesis the two-stage approach has been used because it allowed the analysis of separate outcomes in each trials at the first stage, reducing the IPD to aggregate data (AD). Furthermore having the results from each single study two-stage approach allowed us to calculate the w/s correlation at the first stage needed to fit the model at the second stage of the two-stage IPD-MA.

1.6 Rheumatoid arthritis

Rheumatoid arthritis is one of the most prevalent chronic inflammatory diseases. It primarily involves the joints, but should be considered a syndrome that includes extra-articular manifestations, such as rheumatoid nodules, pulmonary involvement or vasculitis, and systemic comorbidities [21]. Rheumatoid arthritis is the medical condition that motivates the statistical analyses performed in this thesis and therefore, by way of introduction, some background is now provided on the condition itself.

RA affects the immune system, which normally fights infection, attacking the lining of the joints. As a disease, it is characterised by persistent synovitis, systemic inflammation, and

autoantibodies (particularly to rheumatoid factor and citrullinated peptide) [22]. This makes joints swollen, stiff and painful. The small joints of the hands and feet are usually affected first.

1.6.1 Epidemiology

Findings of population-based studies show RA affects between 0.5% and 1.0% of adults in developed countries. The disease is three times more frequent in women than men [22]. The prevalence of RA rises with age and is highest in women older than 65 years, suggesting hormonal factors could have a pathogenic role [23]. Prevalence of RA varies geographically [22]. The disease is common in northern Europe and North America compared with parts of the developing world, such as rural West Africa [23]. These variations are indicative of different genetic risks and environmental exposures. Some evidence suggests incidence of RA might be declining, with onset happening later in life [22].

1.6.2 Management

Several national and regional guidelines for management of RA exist, including recommendations from the American College of Rheumatology (ACR) [24], European League Against Rheumatism (EULAR) [25], and the UK's National Institute for Health and Clinical Excellence (NICE) [22].

- *Treatment of symptoms.* Analgesics reduce pain, and non-steroidal anti-inflammatory drugs (NSAIDs) lessen pain and stiffness. Both groups of drugs are used widely to control symptoms of RA. Evidence for use of analgesics is modest but uncontroversial; support for use of NSAIDs is considerably stronger [22]. However, NSAIDs have lost their historical role as first-line treatment because of concerns about their limited effectiveness, inability to modify the long-term course of disease, and gastrointestinal and cardiac toxic effects [22].

- *Disease-modifying antirheumatic drugs.* DMARDs are a heterogeneous collection of agents for the treatment of RA [22]. They reduce joint swelling and pain, decrease acute-phase markers, limit progressive joint damage and improve function. Methotrexate (MTX) is the dominant DMARD. Sulfasalazine and Leflunomide are also widely used. Their efficacy has been established in placebo-controlled trials [22]. Adverse effects of DMARDs include those that are minor (e.g., nausea) and serious (e.g., hepatotoxicity, blood dyscrasias, and interstitial lung disease) [22].

- *Biological agents.* Biological agents are another class of treatments used for RA therapy. TNF inhibitors were the first licensed biological agents, followed by Abatacept, Rituximab, and Tocilizumab: they are highly effective [22]. The efficacy of biological agents is most obvious in short-term studies in late disease, when placebo responses are low; it is generally less clear-cut in early disease, when active comparators can achieve good responses. Biological agents are combined conventionally with MTX. Adverse events include span reactions and infections at infusion and injection sites. Concerns have also been raised about demyelination and cancer [22].

- *Glucocorticoids.* Short-term glucocorticoids reduce synovitis. In the long term, they decrease joint damage [22] but incur substantial adverse risks, such as infections and osteoporosis, and their overall risk/benefit ratio is deemed unfavourable [22]. However, they can be especially useful in two settings. First, short-term use during flare-ups in disease can lead to rapid improvement and allow other treatments, such as DMARDs, which have a slower onset of action, to be adjusted. Use of steroids in this way is low risk. Oral or intramuscular glucocorticoids are administered by many centres in this setting. Second, intra-articular glucocorticoids are a highly effective local treatment for individual active joints [22].

- *New treatments*. New biological agents in development include drugs that target proximal effects on the immune response and growth factors for T-cell subsets (such as interleukin-17) [16]. New conventional drugs with DMARD-like properties might also

have important future roles. Clinical trials of inhibitors of the kinases have provided promising data, and other targets are under investigation [22].

There are also non-pharmacological treatments such as supportive treatment, for example, effective non-drug treatments, joint protection, foot care and psychological support [22]. Patients' education is also of crucial importance. All these strategies are best delivered by a multidisciplinary team of rheumatologists, nurses, therapists, and podiatrists. Management of comorbidities is important, as they reflect both the disease process and its treatment. Comorbidities include cardiac disease, bone disease and depression. Surgical treatment, particularly joint replacement surgery, is vital to maintain function when joints fail, and collaboration with orthopaedic specialists is required [22].

1.7 Core outcome set

1.7.1 General definition and application of core outcome set

In an RCT, the effectiveness of an intervention is usually determined by comparing outcomes that reflect beneficial and harmful effects. Selection of appropriate outcomes or domains is crucial when designing clinical trials to compare directly the effects of different interventions in ways that minimise bias [26].

There is a growing recognition that insufficient attention has been paid to the outcomes measured in clinical trials [26]. Difficulties caused by heterogeneity in outcome measurement are well known to systematic reviewers. Empirical research provides strong evidence that outcome reporting bias (ORB), defined as the results-based selection for publication of a subset of the original measured outcome variables, is an important problem in randomised trials [27] that affects the conclusions in a substantial proportion of Cochrane Reviews [28]. ORB is likely to affect SRs more widely, as well as published research in general.

These issues could be addressed through the development and use of an agreed standardised collection of outcomes, known as a core outcome set (COS), which should be

measured and reported in all trials for a specific clinical area [29]. It is not compulsory that outcomes in a particular trial should be restricted to those in the COS but there is an expectation that the core outcomes will always be collected and reported, and that researchers will continue to explore other outcomes. The central idea behind the COS is that they should be reflective of the outcomes important to both health care professionals and patients. Examples exist where patients identified an outcome important to them as a group that might not have been considered by practitioners on their own [26]. It should be remembered here that the importance of incorporating health service user opinion in COS development is increasing, but involvement has been limited to date.

The Core Outcome Measures in Effectiveness Trials (COMET) initiative [26] brings together researchers interested in the development, application and promotion of COS, derived using rigorous consensus methods, for effectiveness trials. The objectives of COMET are to collect and stimulate the development of relevant resources, both applied and methodological, to facilitate exchange of ideas and information, to work with patients, the public and their representatives, to develop material to improve health service user engagement, and to increase methodological research in the area of COS [30]. Information on relevant individual studies, both published and ongoing, are being included in a free, publically available internet-based resource. This is a unique resource, which is updated periodically, and which should serve to minimise duplication of effort in the development of COS.

The COMET initiative provides a focus for the continued development of a framework for outcome measurement, first in relation to domains and outcomes within domains, subsequently in terms of definitions and measurement instruments, and finally in relation to the timing of measurement. There is an increasing awareness of the need for greater attention to be given to the outcomes measured in clinical trials, in terms of standardisation and reporting [26].

1.7.2 Core outcome set for rheumatoid arthritis

One of the most remarkable works regarding outcome standardisation has been in the area of RA. The Outcome Measures for Rheumatology Clinical Trials (OMERACT) collaboration was established in 1992; it advocates the use of COS in clinical trials in rheumatology [30]. In 1994, the OMERACT group ratified one of the first COS for RA [8]: tender joint count (TJC), swollen joint count (SJC), pain, physician global assessment (Phy.Global), patient global assessment (Pat.Global), function and acute phase reactant (APR). In studies lasting at least one year, an additional recommendation was for radiographs of the joints to be taken to assess radiological damage (RD).

Previous work conducted by other researchers analysed a review of 350 trials for the treatment of RA of rheumatology trials published up to 2009 [31] identified through The Cochrane Library. Findings of this research were particularly interesting. Indeed, it was discovered firstly that between 60% and 70% of trialists conducting trials in RA measured the core outcomes, and secondly that 90% of trialists contacted said they would consider measuring the RA COS if they were to lead a new trial in RA. However, the fact that it is established that an outcome has been measured does not imply that it has been appropriately reported in the publication of the study. Furthermore, composite outcomes are often reported in place of the individual core outcomes; examples include Disease Activity Score (DAS) [32] and ACR criteria [23]. The CMSG ratifies the use of the COS in its reviews. The consequence of this type of reporting is that many meta-analyses of the individual core outcomes will contain missing data, when it is known that the individual outcomes were measured and possibly analysed. As part of this thesis, these missing data are investigated and it is determined whether or not they impact on the review meta-analyses.

1.8 Thesis structure

ORB is a missing data problem and is a threat to the validity of SRs. There are a number of statistical approaches for adjusting for ORB in meta-analyses but this thesis focuses on the use of multivariate meta-analysis (MVMA). MVMA allows the joint synthesis of outcomes

simultaneously and utilises correlations across studies which has the potential to reduce the impact of ORB. The aims of this thesis were two. The first objective was to apply the methods to assess the presence of outcome reporting bias with multiple outcomes. The second objective was to understand the use of MVMA through simulation studies and application of the method to reviews of RA, in order to improve statistical analysis methods for MA research, especially when considering the risk of bias of the available evidence. In **Chapter Two** a framework for assessing ORB in SRs is presented and the methods applied to a cohort of Cochrane reviews in RA using the core set of outcomes for this condition. The prevalence of high risk ORB in the trials included within these reviews is estimated. **Chapter Three** presents an IPD FE MA simulation study to compare the performance of the MVMA method against univariate meta-analysis (UVMA) under a range of different study-level missing data (ORB) scenarios. **Chapter Four** presents a similar simulation study to compare the performance of the MVMA method against UVMA when there is missing individual *patient-level* missing data (missing data in the IPD). In **Chapter Five**, MVMA is applied to the RA examples where ORB was assessed in Chapter 2. The results are compared to univariate approaches and the advantages and disadvantages of each approach are discussed. The impact the missing data have on the original MA results in term of the statistical conclusions is also quantified and discussed. **Chapter Six** contains a discussion, conclusions and recommendations for further work. This research is the first to assess the prevalence of ORB against a set of core correlated outcomes. It also demonstrates the feasibility of using a MVMA approach to examine the robustness of MA results when there is high-suspicion of ORB. Finally, the real novel aspect of this work is that, under the generation of IPD, we can also explore the benefit MVMA might have on missingness at the patient level, which cannot be investigated in AD MA.

Chapter 2 – Outcome reporting bias

2.1 Background

In Chapter 1, the fundamental concepts of EBM, clinical trials, MA and SRs were introduced.

From the literature, it is clear that different types of bias could affect the results of a MA, particularly when outcome data are missing [34, 35]. Study publication bias (where whole studies are not published) and w/s selective outcome reporting (where a subset of the originally recorded outcomes has been selectively reported) have been recognised as two missing data problems that can affect EBM. If selective reporting or non-reporting is driven by the significance and/or direction of the effect size (for example non-significant outcomes are reported only as p-value > 0.05 or are suppressed altogether) then this is known as selective reporting bias or ORB [36].

Empirical evidence suggests that there is strong evidence of an association between significant results and publication; studies that report positive or significant results are more likely to be published and outcomes that are statistically significant have higher odds of being fully reported (FR) (range of odds ratios: 2.2 to 4.7) [37]. In this chapter the focus is on ORB, although it is important to note that study publication bias and ORB often impact review meta-analyses in the same way. Impact assessment is addressed in Chapter 6. A recent study has suggested that the amount of missing participant data from unpublished studies (15%) is greater than that from published studies (4%) [33]. However, if partially reported data is considered to be unpublished then the amount of missing data from published studies increases to 38%. Considering partially reported data as missing is an important consideration since it reflects the fact that the data cannot be included in a review MA. Furthermore, partial reporting has previously been associated with a high risk of ORB [37].

Results from the Outcome Reporting Bias in Trials (ORBIT) study [28] have shown that over half of SRs did not include full data for a single review primary outcome of interest from all eligible trials and half of the trials assessed with missing data were under high suspicion of ORB [37].

One way to reduce the problem of ORB is the introduction of an agreed minimum set of standardised outcomes, to be measured and reported in all trials for a particular disease or condition, referred to as a COS [30]. In Chapter 1, a COS for RA which was defined by OMERACT in 1994 [28] was introduced. Only one previous study has assessed ORB for all review outcomes in cystic fibrosis [38]. This study showed that all eligible trials from all published reviews from the Cochrane Cystic Fibrosis group did not include full data when looking across all review outcomes; that is, both primary and secondary outcomes [38].

In these SRs a core set of outcomes has been endorsed by trialists and the Cochrane Review Group. While uptake of this COS has been shown to be improving over time (up to 70% of trialists measuring the full set of core outcomes in 2010), *reporting* of the outcomes appeared to be less well performed, making the studies prone to potential ORB [27].

The aim of this chapter is to present the methods to determine how to detect and assess suspected ORB in SRs and to estimate the prevalence of ORB in the Cochrane RA reviews by considering all eight outcomes in the COS.

2.2 How to assess outcome reporting bias in a systematic review

A tutorial for assessing the potential for ORB in a review exists [28], although the tools available, particularly the impact methods (to be addressed in Chapter 6) to systematic reviewers to help them complete the assessments have largely been updated since the tutorial was published. In this section, a review of the most recent methodologies and tools available for assessing ORB in reviews is presented in a series of steps. The methods require the use of two important tools, an outcome matrix for detecting missing data and a classification system for assessing the potential risk of ORB.

This methodology has already been successfully implemented by some systematic reviewers [40].

2.2.1 Steps for addressing missing outcome data in a review

Step 1: Exclusion criteria

The first step is to ensure that no potentially eligible studies are excluded from the review for the sole reason of not reporting on any review outcomes of interest. If a study report does not give results for or mention particular outcomes, this does not necessarily mean that they were not measured or analysed. For this reason, studies must not be excluded if they do not report any of the relevant review outcomes. For example, the review 'Methotrexate monotherapy versus methotrexate combination therapy with non-biologic disease modifying antirheumatic drugs for RA' [41] excluded from the assessment the study Mottaghi 2005 [42] and the reason for exclusions was that this trial contained "no outcome of interest".

Step 2: Constructing the outcome matrix from the study reports

The outcome matrix is constructed by listing all the eligible studies as rows and all the review outcomes of interest as columns in the matrix. Outcomes can be distinguished in terms of review primary and secondary outcomes and as benefit or harm outcomes. Outcomes that are not of interest in the review but are reported in the reports for eligible trials are also listed. This has been found to be useful when assigning the risk of bias. For example, in RA, composite disease activity criteria are often reported, e.g. ACR response criteria. The ACR criteria [43] considers TJC, SJC and three of the following five outcomes: patient's assessment of pain, patient's global assessment, and Phy.Global, patient's assessment of physical function and acute-phase reactant. For this reason, if ACR criteria are reported then it is known that all of the RA core outcomes (with the exception of RD) have been measured, even if they are not reported. Furthermore, some outcomes are often routinely measured together so that, if one outcome is reported but not the other, this may

raise suspicion that selective reporting has occurred, e.g. TJC and SJC. A tool for constructing the outcome matrix (ORBIT matrix generator) is freely available (http://ctrc.liv.ac.uk/orbit/) [28].

Step 3: Completing the outcome matrix

Once the outcome matrix has been constructed, it can be filled in (the matrix can also be filled in using the ORBIT matrix generator) [28]. For each study, a reviewer should indicate which outcomes (benefits and harms) were reported and differentiate between 'full reporting', 'partial reporting', 'not reported – not clear whether measured or not' and 'not measured'. Guidance on what constitutes the different levels of reporting can be found below.

a) *Full Reporting*. Studies that 'fully report' an outcome should be denoted by a tick in the matrix. Typically, if a study includes enough information on the outcome to be included in a review MA (without contact for extra information from the trialists), then the outcome is FR and there is no risk of non-reporting bias. If a MA is not appropriate, then the outcome can still be considered FR. Some guidance on what constitutes full reporting for each of the different outcome types (e.g. binary, continuous and time to event) is provided in Box 1 below.

Box 1. Data Required for meta-analysis of Fully Reported Outcomes [44]

For Unpaired Continuous Data

Sample size in each group and magnitude of treatment effect (group means/medians or difference in means/medians) and measure of precision or variability (confidence interval, standard deviation, or SE for means; interquartile or other range for medians) or the precise *p-value**

For Unpaired Binary Data

Sample size in each group and either the numbers (or percentages) of participants with the event for each group, or the odds ratio or relative risk with a measure of precision or variability (confidence interval, standard deviation, or SE) or the precise *p-value**

For Paired Continuous Data

Sample size in each group and either the raw data for each participant, or the mean difference (MD) between groups and a measure of its precision or variability or the precise *p-value**

For Paired Binary Data

Sample size in each group and paired numbers of participants with and without events

For Survival Data

Either a Kaplan-Meier curve or similar, with numbers of patients at risk over time, or a hazard ratio with a measure of precision and sample size in each group

*Sample sizes, treatment effect and precise *p-value* enable the calculation of a SE if a measure of precision or variability is not reported.

b) *Partial Reporting*. Studies that 'partially report' an outcome should be denoted by an open circle in the matrix. For partial reporting, it is clear that the review outcome of interest has been analysed in the eligible studies but there is not enough information reported for the outcome to be considered FR. Some suggestions for partially reported outcomes include:

- p-value reported only (with no treatment-effect size indicated) (high risk of bias)
- No measure of variance/precision reported (low risk of bias)
- Percentage data given for both groups but denominators unclear (binary data) (low risk of bias)
- Results are reported graphically only (accurate data extraction not possible) (low risk of bias)
- Treatment-effect estimate reported from a statistical model only (low risk of bias) (this would be accepted as full reporting if the SE of the estimate is also given).

c) *Not reported.* This means that it is not clear whether the outcome has been measured or not. If a study does not report on a review outcome, it should be denoted with a cross in the matrix. Reviewers should compare all sections of the study reports to the results section, by looking at which other outcomes were measured and reported and accounting for knowledge of the clinical area. Reviewers should be suspicious of high risk of bias, if it is either clear or assumed that the outcome had been measured and possible that non-reporting could have been influenced by the results.

d) *Not measured*. This means that in some situations it will be clear from the study reports that outcomes were not measured. In these situations, studies should be denoted with a star in the matrix. If it is clear from the study report that outcomes were not

measured, then this eliminates any risk of non-reporting bias. For example, in RA, studies that last for less than one year often do not measure RD (as damage progression is long term). If a study therefore reported “radiographic assessments were not made in this study” then it is clear that RD was not measured.

Step 4: Contacting trialists

After the outcome matrix has been completed, reviewers should make an attempt to contact the trialists from the studies included in the review that partially reported the review outcomes of interest or where it was not clear whether the outcome was measured or not. The purpose of this contact is to try and obtain missing outcome data to include in the review analysis or to confirm that the outcomes of interest were not measured. The matrix should be updated accordingly in light of the extra information obtained.

Step 5: Outcome reporting bias assessment: classification system

Once the outcome matrix is complete and trialists have been contacted for missing data, a reviewer should then assess the potential risk of ORB as a result of outcomes being partially reported or not reported using the ORBIT nine-point classification system for benefit outcomes (Table 2.1) [28]. Each outcome with missing or partially reported data should be given a classification. Each classification should be justified with a reason, using verbatim study report text whenever appropriate. The classification is determined by comparing all sections of the study reports to the results section of the trial report, by looking at which other outcomes were measured and reported and accounting for knowledge of the clinical area. As with data extraction and risk of bias assessment, this task should be completed by at least two researchers independently and differences should be discussed to agree on an overall classification for each review outcome that is missing for each study. If available, review authors should also seek to obtain study protocols or trial registry entries for eligible studies in order to compare pre-specified outcomes with those reported in the final study report. The comparison between what was planned (in terms of outcome specification) and what was actually measured has the potential to simplify the risk of bias assessment for

each outcome. The classification system was developed to assess the risk of bias when a trial was excluded from a MA either because the data for the outcome were not reported or because the data were reported incompletely (for example, just as “non-significant”). The categories reflect the stages of assessing whether an outcome was measured, whether an outcome was analysed, and, finally, the nature of the results presented. The system identifies whether there is evidence that the outcome was measured and analysed but only partially reported (A to C classifications), whether the outcome was measure and analysed but not reported at all (D classification), whether the outcome was measured but not necessarily analysed (E and F), if it is unclear whether the outcome was measured (G and H), or if it is clear that the outcome was not measured (I).

- A “*high risk*” classification is assigned when it is either known or suspected that the results were partially or not reported because the treatment comparison for the outcome of interest was statistically non-significant (p-value > 0.05).

- A “*low risk*” classification is set when it is suspected but not actually known that the outcome was either not measured, measured but not analysed, or measured and analysed but either partially reported or not reported for a reason unrelated to the results obtained.

- A “*no risk*” classification is reserved for cases where it is known that the outcome was not measured, known that it was measured but not analysed, or known that it was measured and analysed but the reason for partial or no reporting is not because the results were statistically significant.

  For the cases where the outcome was measured but not necessarily analysed, judgment is needed as to whether it was likely (E) or unlikely (F) that the measured outcome was analysed and not reported because of non-significant results. When it is unclear whether the outcome was measured, judgment is also needed as to whether it is likely that the outcome was measured and analysed but not reported on the basis of non-significant results (G) or unlikely that the outcome was measured at

all (H). In the original ORBIT study [28], two separate sensitivity and specificity analyses were performed. The first analysis considered only G and H classifications and aimed to determine how good the classification system was at judging whether the primary outcome of interest in the review had been measured when it was not mentioned in the trial report. The second analysis looked at how well the bias could be predicted when a judgment was required. The truth was obtained by contacting the trialist to determine whether the outcome was measured or not and, if it was measured, was the outcome analysed, and what was the justification for non-reporting. The sensitivity and specificity results were high, demonstrating that the classification system was able to distinguish satisfactorily between an outcome being measured or not and being able to predict bias reliably.

Trial classified as A/D/E/G, C/F/H, and B/I are assumed to be at high, low and no risk of ORB, respectively, in relation to the review outcomes. In a similar way, all missing review outcomes from trials excluded from the review but selected for assessment should also be assigned an ORBIT classification [28].

Table 2.1 ORBIT study classification for benefit outcomes [28]

	Description	Level of reporting	Risk of bias
	Clear that the outcome was measured and analysed		
A	*Trial report states that outcome was analysed but only reports that result was not significant (typically stating p-value > 0.05).*	*Partial*	*High Risk*
B	*Trial report states that outcome was analysed but only reports that result was significant (typically stating p-value < 0.05).*	*Partial*	*No Risk*
C	*Trial report states that outcome was analysed but insufficient data were presented for the trial to be included in meta-analysis or to be considered to be fully tabulated.*	*Partial*	*Low Risk*
D	*Trial report states that outcome was analysed but no results reported.*	*None*	*High Risk*
	Clear that the outcome was measured		
E	*Clear that the outcome was measured. Judgment says outcome **likely** to have been analysed but not reported because of non-significant results.*	*None*	*High Risk*
F	*Clear that the outcome was measured. Judgment says outcome **unlikely** to have been analysed.*	*None*	*Low Risk*
	Unclear whether the outcome was measured		
G	*Not mentioned but clinical judgment says likely to have been measured and analysed but not reported on the basis of non-significant results.*	*None*	*High Risk*
H	*Not mentioned but clinical judgment says unlikely to have been measured at all.*	*None*	*Low Risk*
	Clear that the outcome was not measured		
I	*Clear that the outcome was not measured.*	*NA*	*No Risk*

***Risk of bias** arising from the lack of inclusion of non-significant results when a trial was excluded from a meta-analysis or non-fully reported in a review because the data were unavailable.*

2.3 Assessment of outcome reporting bias in Cochrane systematic reviews of rheumatoid arthritis

A cohort of SRs published by the CMSG (up to and including the September 2012 issue) that considered pharmacological interventions (DMARDs, biologics or glucocorticoids) for the treatment of RA were included. Reviews were identified via the Cochrane topics link (http://onlinelibrary.wiley.com/book/10.1002/14651858/topics) and were those that were indexed under 'musculoskeletal', RA, 'treatment (pharmacological interventions)' and 'biologics/steroids/DMARDs'. The reviews were selected based on previous work conducted by Kirkham et al. [39]. Furthermore this assessment was focused on pharmacological interventions because the scope of the COS [40] was not specifically designed for non-drug trials and does not focus on measures of safety. Therefore systematic reviews that

considered non-pharmacological interventions or considered drug safety only were excluded from this work of assessment. I decided to focus on systematic reviews up to 2012 because the Cochrane SRs set was the most updated list meeting the eligibility criteria to be included in the assessment of ORB.

Overviews and reviews that contained no eligible RCTs (empty reviews) were also excluded, as an assessment of primary studies would not be possible. Empty reviews were however checked to make sure no studies were excluded due to 'no relevant outcome data' (see Step 1, Section 2,1,1) which could be included in the assessment. Reviews were assessed for ORB in relation to eight outcomes for RA. In some cases, it was often problematic to assess whether an outcome was measured and clinical judgment was required. If there were any uncertainties with the classifications, review authors were consulted. All trials were independently classified by two researchers (G.F. and J.J.K.), and disagreements were resolved through discussion. The full assessment followed the procedures outlined in Section 2.2. Trialists were not contacted for missing outcome data in this instance (Step 4), as this had already been carried out by the authors of the included reviews. Such outcomes have not previously been addressed when it comes to ORB assessments. Firstly, the purpose is to discuss how these outcomes were dealt with in order to provide 'rules' for assessing the risk of bias that could be consistently applied to all trials.

2.3.1 Outcome reporting bias assessment for composite outcomes

Trials of RA often use composite outcomes to measure disease activity. These composite outcomes often encompass a number of the core individual outcome measures. For instance, in section 2.2.1 we explained in detail the composite measure called ACR, but others also exist, for example, the DAS [45]. The DAS is a composite index that includes the combination of the values of TJC and SJC, patient's global assessment of disease activity, and APR measures as erythrocyte sedimentation rate (ESR) value. When composite outcome criteria were reported in full, but no data on any of the individual core outcomes were reported, then the low-risk F-classification was used (clear that the outcome was measured but unlikely that the individual outcomes were analysed) for all core outcomes

contained within the composite. The motivation is that it may not have been the trialists' intention to analyse the individual core outcomes. If a trialist selectively reported some of the individual core outcomes from the composite, then the high-risk E-classification was used for the core outcomes not reported, as in this situation it is more likely that all the core outcomes would have been analysed, and the likely reason for some of them not being reported is a non-significant result.

2.4 Results

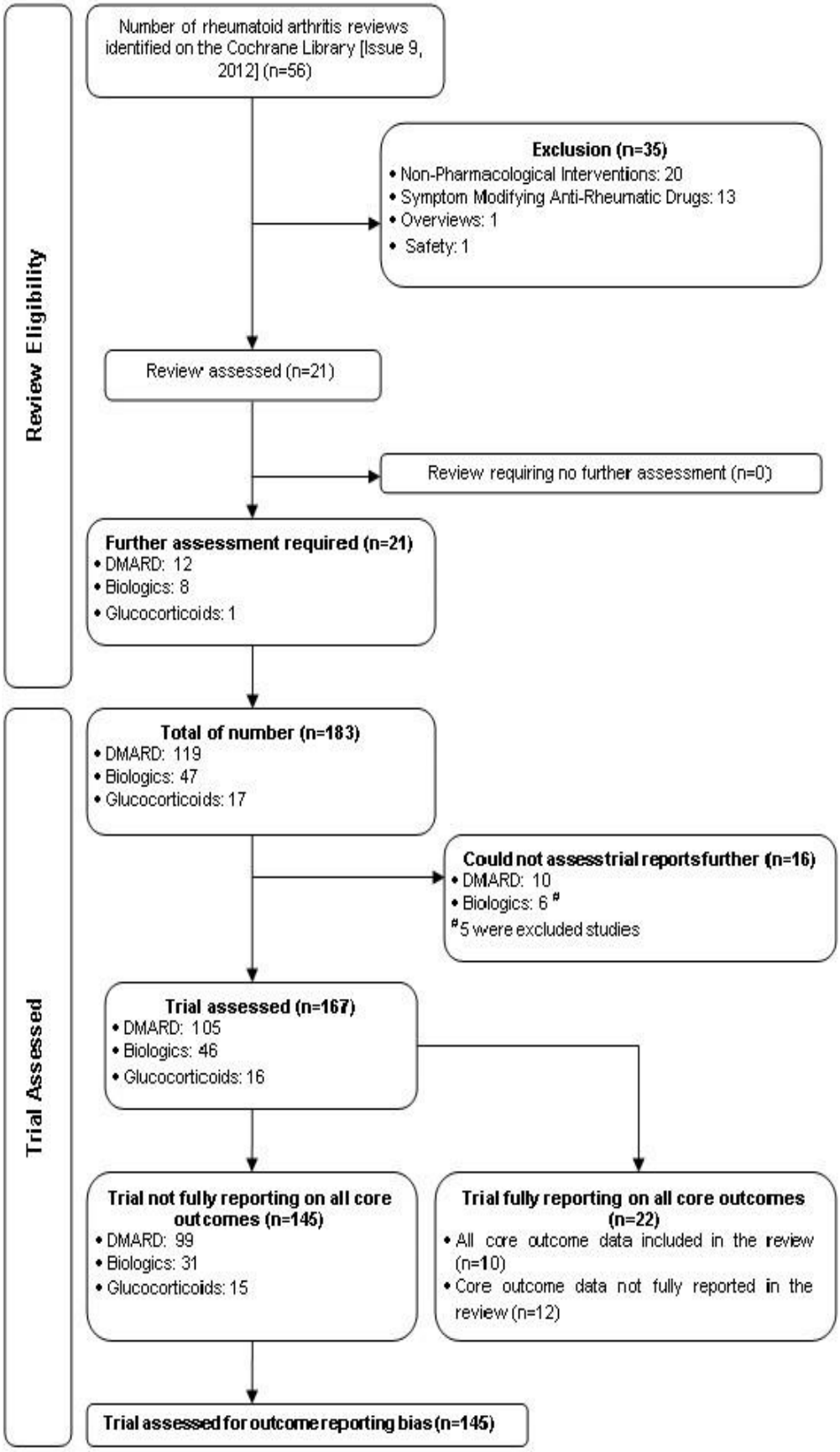
2.4.1 Systematic reviews assessed: eligibility criteria

The CMSG published 56 unique RA reviews up to and including the September 2012 issue (Figure 2.1). Thirty-five reviews were excluded: 20 focused on non-pharmacological interventions, 13 studied symptom-modifying antirheumatic drugs, one was an overview, and one focused on safety.

Of the remaining 21 reviews included in the assessment, 12 reviews considered DMARDs, eight considered biologics, and one considered glucocorticoids. All reviews required an ORB assessment for at least one eligible trial.

The 21 reviews included a total of 183 trials for assessment (Figure 2.1). All the reports for each randomised trial included in the eligible reviews were obtained for evaluation. Among the 183 trials included within the 21 reviews, 16 trials could not be assessed further because either the articles were not in English (n=10) or the trial reports were unobtainable (n=6). Therefore, in this study, 167 trials were assessed, 105 on DMARDs (63%), 46 on biologics (28%) and 16 on glucocorticoids (9%) (Figure 2.1). Of the 167 assessable trials (Figure 2.1), 22 trials (13%) FR the outcome data for all core outcomes, but the data in only ten of these were adequately reported in the review. Among these 22 trials, ten studies included all core outcome data in the review and for twelve studies the core outcome data were reported only in the study but not completely FR in the review.

Figure 2.1 Flow diagram of reviews eligibility criteria and assessment of randomised controlled trials within reviews



A summary of the studies assessed in each of the 21 SRs is provided in Table 2.2.

Table 2.2 Summary of the 167 RCTs assessed in the 21 SRs

Systematic Review	Objective	Number of trials not fully reporting on all core outcomes		Number of trials fully reporting on all core outcomes
		Included	Excluded	
Methotrexate monotherapy versus methotrexate combination therapy [41]	Efficacy and toxicity	19	2	0
Antimalarials [46]	Efficacy and toxicity	4	2	0
Azathioprine [47]	Short-term effects	3	3	0
Auranofin [48]	Efficacy and toxicity	9	1	0
Cyclophosphamide [49]	Short-term effects	2	0	0
Cyclosporine [50]	Short-term (up to one year) effects	2	0	1
Injectable gold [51]	Short-term benefit and risk of side-effects	3	0	1
Methotrexate [52]	Short term efficacy and toxicity	5	0	0
Penicillamine [53]	Short-term effects	7	0	0
Sulfasalazine [54]	Short-term efficacy and toxicity	7	0	0
Leflunomide [55]	Efficacy and toxicity	20	1	4
Folic acid and folinic acid for reducing side effects in patients receiving methotrexate [56]	Effects of folic acid and folinic acid in reducing the mucosal and gastrointestinal (**GI**) and haematologic side effects of low-dose of MTX in patients with RA and to determine whether or not folate supplementation alters MTX efficacy	7	2	0
Adalimumab [57]	Efficacy and safety	3	0	3
Abatacept [58]	Efficacy and safety	5	0	2
Anakinra [59]	Effectiveness and safety	3	0	2
Certolizumab pegol [60]	Effectiveness and safety	3	0	1
Etanercept [61]	Update the previous Cochrane systematic review published in 2003 assessing the benefits and harms	6	0	3
Infliximab [62]	Efficacy and safety	1	0	1
Golimumab [63]	Efficacy and safety	2	0	2
Tocilizumab [64]	Efficacy and safety	7	1	1
Glucocorticoids [65]	Evaluating efficacy in inhibiting the progression of RD in RA	15	0	1
		133	**12**	**22**
		Total: 145		

Table 2.2 shows that, not taking into account studies that are fully reporting on all core outcomes, the total number of RCTs not fully reporting on all core outcomes that have been assessed in these SRs for ORB was 145.

This set of studies contains 133 (92%) trials that were included in the SR. Trials that were excluded in the 'characteristics of excluded studies' section were also checked for any suggestion of ORB. The number of excluded studies was 12 (8%) (see Table 2.2).

2.4.2 Constructing and completing the ORBIT matrix

Using the information from the reviews, the ORBIT matrices [28] were constructed for each review. As an example, the matrix is provided for the review of Auranofin [48] for the treatment of RA (Table 2.3). In this particular review, none of the included trials reported on the full set of core outcomes. The matrix also identifies that no trial reported on RD and Borg 1991 reported on none of the core outcomes. Borg 1991 was excluded from the original SR because no OMERACT outcomes were reported in this article. In the appendix, the ORBIT matrices are shown for the remaining 20 SRs assessed (Appendix A, Tables A.1-A.20).

Table 2.3 Constructing the ORBIT matrix - Auranofin for treating rheumatoid arthritis [48]

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Davies '82	Included	✓	×	✓	×	×	×	×	×
Prouse '82		✓	×	✓	×	×	×	✓	×
Palmer '82		✓	✓	×	✓	×	✓	✓	×
Ward '83		✓	✓	✓	✓	✓	×	✓	×
Wenger '83		✓	✓	✓	×	✓	×	✓	×
Lewis '84		×	×	×	×	×	×	✓	×
Bombardier '86		✓	✓	✓	✓	✓	✓	✓	×
Johnsen '89		✓	✓	✓	✓	×	×	✓	×
Glennas '97		×	✓	✓	×	×	✓	×	×
Borg '91	Excluded	×	×	×	×	×	×	×	×

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage

2.4.3 Applying the ORBIT classification system: Auranofin for treating rheumatoid arthritis

Table 2.4 shows the outcome matrix (Auranofin for treating RA [48]), which was constructed using the outcome data that was included in the SR. The sample size for all participants in this review was 1287, ranging from 20 patients in the Palmer 1982 trial to 340 in the Wenger 1983 trial.

Table 2.4 Completing the ORBIT matrix - Auranofin for treating rheumatoid arthritis [48]

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Davies '82		✓	×(H)	✓	×(H)	×(H)	×(C)	×(H)	×(D)
Prouse '82		✓	×(H)	✓	×(H)	×(H)	×(C)	✓	×(E)
Palmer '82		✓	✓	×(C)	✓	×(H)	✓	✓	×(H)
Ward '83		✓	✓	✓	✓	✓	×(FR)	✓	×(H)
Wenger '83	Included	✓	✓	✓	×(H)	✓	×(C)	✓	×(FR)
Lewis '84		×(F)	×(H)	×(E)	×(H)	×(H)	×(F)	✓	×(H)
Bombardier '86		✓	✓	✓	✓	✓	✓	✓	×(H)
Johnsen '89		✓	✓	✓	✓	×(D)	×(FR)	✓	×(FR)
Glennas '97		×(G)	✓	✓	×(H)	×(H)	✓	× (A)	×(FR)
Borg '91	Excluded	×(G)	×(D)	×(D)	×(D)	×(G)	×(C)	×(G)	×(G)

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: Acute Phase Reactant; RD: Radiological Damage; FR Fully Reported

In Table 2.4, eight core outcomes for the assessment of the efficacy of RA are presented, TJC, SJC, pain, Pat.Global, Phy.Global, function, APR and, finally, RD.

As all the trials had missing data on at least one outcome; all trials were scrutinised and the outcome classified using the ORBIT classification system [28] where an outcome was not reported in the review. Table 2.4 shows the classifications for the Auranofin review [48], taking into account the information from the trial report. This table shows that, for 48% (38) of the 80 evaluable outcomes (8 outcomes x 10 trials) carried out in this SR, the set of core outcomes was either partially reported or not reported (A to I classification). For 16%, 13 of the 80 assessments carried out in this SR, at least one core outcome was classified under high suspicion for ORB (A, D, E, or G classification), whereas, for 14% (11 of 80), it was

clear the outcomes were measured and analysed (A, B, C, D classification), but the reporting of the outcomes meant that the data could not be included in a MA. Some trials did report data on function and RD but they were not considered when calculating the pooled results of the MA. Notably, the outcome function for the Ward [66] and Johnsen [67] trials was classified as 'FR'. Ward measured, analysed and reported in full the outcome grip strength, which was not a validated measure of function for this particular review. However, amongst rheumatologists now this method to measure the outcome function is considered a consistent measure and it is accepted and therefore function was classified as FR also when it has been measured with grip strength. Johnsen on the other hand measured, analysed and reported in full a 'Health Assessment Questionnaire (HAQ)' but it was not reported in the MA. The HAQ was a validated measure of function used in the review; this outcome was therefore classed as FR and this was missed by the reviewers. Furthermore, RD for the Wenger, Johnsen and Glennas trials was classified as 'FR'. In detail, Wenger did not use a validated scale to measure RD but Johnsen and Glennas used 'Larsen Dale index' reporting respectively median and Inter Quartile Range (IQR) and median (range) and p-value. While this non-parametric data could not have easily been used in the review MA, the reviewer should have tabulated this data within their review.

Table 2.5 provides the justification for these classifications. The classifications and the remaining 20 reviews can be found in the appendix to this chapter (Appendix A).

Table 2.5 Justifications for the ORBIT classifications: Auranofin for treating RA [48]

Trial	Outcome	Classification	Reason
Davies '82	SJC	H**	SJC is not mentioned in this trial
	Pat.Global	H**	Pat.Global is not mentioned in this trial
	Phy.Global	H**	Phy.Global is not mentioned in this trial
	Function	C	Grip strength was partially reported (mean only)
	APR	H**	APR is not mentioned in this trial
	RD	D	RD clearly measured but no results were reported
Prouse '82	SJC	H**	SJC is not mentioned in this trial
	Pat.Global	H**	Pat.Global is not mentioned in this trial
	Phy.Global	H**	Phy.Global is not mentioned in this trial
	Function	C	Grip strength was partially reported (graphically)
	RD	E**	RD measured and likely analysed – composite and a selective set of outcomes reported
Palmer '82	Pain	C	Pain is reported only graphically
	Phy.Global	H**	Phy.Global is not mentioned in this trial
	RD	H**	This study had a duration of 6 months
Ward '83	Function	FR	The trialists measured, analysed and reported in full functional class
	RD	H**	This study had a duration of 21 weeks
Wenger '83	Pat.Global	H**	Pat.Global is not mentioned in this trial
	Function	C	Grip strength was only partially reported (MD only)
	RD	FR	n/N progressed + radiological score changes reported
Lewis '84	TJC	F**	TJC measured but unlikely analysed
	SJC	H**	SJC is not mentioned in this trial
	Pain	E**	Outcome measured and likely analysed – composite and a selective set of outcomes reported
	Pat.Global	H**	Pat.Global is not mentioned in this trial
	Phy.Global	H**	Phy.Global is not mentioned in this trial
	Function	F**	Function measured but unlikely analysed
	RD	H**	This study had a duration of 6 months
Bombardier '86	RD	H**	This study had a duration of 6 months
Johnsen '89	Phy.Global	D	Phy.Global measured but no results were reported
	Function	FR	The trialists measured and reported functional class
	RD	FR	Reported as median (IQR) for both treatment groups
Glennas '97	TJC	G**	Routine practice to measure TJC
	Pat.Global	H**	Pat.Global is not mentioned in this trial
	Phy.Global	H**	Phy.Global is not mentioned in this trial
	APR	A	Outcome measured and analysed: p-value reported only as p>0.05
	RD	FR	Trialists measured and reported functional class
Borg '91	TJC	G**	Routine practice to measure TJC
	SJC	D	SJC clearly measured but no results were reported
	Pain	D	Pain clearly measured but no results were reported
	Pat.Global	D	Pat.Global clearly measured but no results were reported
	Phy.Global	G**	Routine practice to measure Phy.Global
	Function	C	Function is reported only graphically (HAQ)
	APR	G**	Routine practice to measure APR
	RD	G**	Routine practice to measure RD in long-term trials

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: Acute Phase Reactant; RD: Radiological Damage; FR: Fully Reported

**E/F and G/H classifications were confirmed with the reviewer

2.4.4 Assessment results: all rheumatoid arthritis reviews

The ORBIT classification [28] for all the assessed trials is shown in Table 2.5. A breakdown of the classification by intervention class (DMARDs, biologics, and glucocosteroids) is provided in the appendix to this chapter. In the following table (Table 2.6) the clinical trials considered for the assessment of ORB are 145 that are derived from the set of 167 studies except the 22 studies for which the outcomes were FR.

Table 2.6 Clinical trials assessed for outcome reporting bias

Class	Rheumatoid Arthritis Core Set of Outcomes									TOTAL (%)[a]
	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD		
								< 52 wk	≥ 52 wk	
A	3	4	1	3	2	2	4	2	2	21 (1.7%)
B	1	1	0	0	1	0	0	0	0	3 (0.2%)
C	21	19	17	20	18	33	36	1	4	168 (13.9%)
D	2	3	8	8	8	8	7	2	2	46 (3.8%)
E	19	16	18	22	20	8	15	2	0	118 (9.8%)
F	10	10	8	7	6	7	8	0	1	57 (4.7%)
G	7	12	20	15	19	7	4	6	6	90 (7.5%)
H	6	8	19	18	24	10	7	89	0	92 (7.6%)
I	0	0	0	0	0	0	0	4	0	0 (0%)
Total: A to I classification										595 (49.4%)
Fully reported	98	94	76	74	69	92	86	25	21	610 (51%)
Total	167	167	167	167	167	167	167	131	36	1205
TOTAL Missing Data (A-I) (%)[b]	69 (41%)	73 (44%)	91 (54%)	93 (56%)	98 (59%)	75 (45%)	81 (49%)	N/A[c]	15 (42%)[d]	

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage.

[a] The denominator used is 1205. That is the total number of data points if all 145 trials reported on all seven core outcomes (TJC, SJC, Pain, Pat. Global, Phy.Global, Function, APR) plus the 36 trials that should have also measured and reported on RD due to a follow-up greater than 52 weeks (i.e. (167*7) + (36*1) = 1205). The numerator also excludes the assessment of RD for trials less than 52 weeks.

[b] The denominator used is the total number of trials where an assessment is possible (167).

[c] Not applicable: OMERACT recommends outcome only applicable if follow-up > 52 weeks.

[d] The denominator used is the total number of trials where an assessment is possible and the follow-up is greater than 52 weeks (36).

At the trial level, missing or incomplete reporting of outcome data for each core outcome ranged from 41% (69 of 167 trials) for TJC, up to 59% (98 of 167 trials) for Phy.Global. For 595 (49.4%) of the 1205 evaluable outcomes in this current study, the set of core outcomes

was either partially reported or not reported (A to I classification) (Table 2.6). The 23% (275) of the 1205 assessable outcomes were classified under high suspicion for ORB (A, D, E, or G classification), whereas for 20% (238 of 1205) it was clear the outcomes were measured and analysed (A, B, C, D classification), but the reporting of the outcomes meant that the data could not be included in a MA.

2.5 Discussion

This is the first study to consider an assessment of ORB against a well-established core set of outcomes. The OMERACT COS was approved for use in clinical trials for RA but is also endorsed by the CMSG [6]. Although the uptake of the measurement of the COS for rheumatoid arthritis trials has been shown to be increasing [40], the reporting of the outcomes for many of these trials remains insufficient, meaning that many meta-analyses are unable to include data from all relevant studies. Similar to the study reported by Dwan et al [38], who looked at all review outcomes in a cohort of Cochrane Cystic Fibrosis reviews, all the reviews considered in this study included at least one study that contained missing data in relation to at least one of the core outcomes. Across all core outcomes, there were 238 items of study data missing from meta-analyses for outcomes that were clearly measured and analysed (A-D classifications), and a further 208 items of study data where the outcome data were clearly measured (or likely measured) but not reported because of non-significant results (E and G classifications). In this assessment, the number of outcomes with high suspected bias (A, D, E, and G) was 275 of the 1205 evaluable outcomes assessed.

It is important that trialists follow CONSORT 2010 [68] guidance for reporting trial findings. Adherence to CONSORT would ensure that all outcomes are reported in full and all pre-specified outcomes are defined and reported. Many outcomes were not mentioned in the trial reports, meaning that clinical judgment was needed to decide whether the outcome of interest was likely to have been measured for a particular trial. Reviewers were contacted to help with the assessments and to provide their expert judgment in these situations, although a limitation of this study is that trialists were not contacted to confirm whether the outcome

was measured or not. This was a pragmatic choice, as the median publication date for trials included in this review was 1999, meaning there would be difficulties in locating many of the trialists.

The original ORBIT study [28] found that the sensitivity and specificity of determining whether an outcome was measured or not and detecting bias from unpublished/partially reported outcomes was excellent. The work in this chapter however demonstrates that the reliability of SRs could be improved if more attention is paid to missing outcomes from the source trial reports. If data are missing, in the first instance, reviewers should be encouraged to at least attempt to contact the trialists or study sponsors to confirm whether the outcome was measured and analysed, and, if so, obtain the results and update the review MA accordingly with the newly obtained data. If data are not available, then reviewers should consider sensitivity analyses to adjust and assess the impact the biases identified have on the pooled effect estimates – a topic covered in Chapter 3.

Another underlying issue in this study is that there are several measurement scales that can be used to measure each of the core outcomes, some of which may not have been accepted as valid measurement instruments in the individual reviews. The RA COS used in this study addresses the issues of ‘what’ outcomes to measure but not ‘how’ to measure them. As part of the uptake study [10], acceptable measurement instruments for each core outcome were decided in advance by expert rheumatologists. Trials were therefore classed as fully reporting the core outcome if they had reported in full one of these acceptable measurement instruments to represent the core outcome that was not of interest in the review. The motivation here is that, in many cases, it was thought unlikely that trials would use several measurement scales to use the same outcome and therefore the risk of non-reporting bias would be low. In addition, some of the measurement tools required for the review may not have been validated at the time the trial was conducted. However, the risk of bias from reported results, i.e. when an outcome is selectively reported based on a subset of the analysis undertaken (e.g. selectively reporting the most significant scales from a battery of measurements used to measure the same outcome), may remain high. This form of

reporting bias was evident in the reporting of composite outcomes which was addressed in this study. In an ideal situation, if a composite outcome is used, then all of its individual components should also be reported in full for complete transparency.

One final comment is that the RA COS set does not address harm outcomes. This chapter outlines a framework for identifying and assessing ORB in SRs. Using a similar framework, harms can also be assessed for ORB in reviews, but the classification for assessing the reasons for missing data may differ from that for benefit outcomes. For harm outcomes, Saini et al [69] propose a different 13-point classifications to be used where the bias is associated with the suppression (or inappropriate) reporting of specific harms data that mask the harm profile of particular interventions.

Chapter 3 - Multivariate fixed-effects meta-analysis to adjust for outcome reporting bias: a simulation study

3.1 Background

In Chapter 2 the notion of ORB was defined, describing the potential impact this form of bias may have on a SR. Furthermore, the results of the study were presented where ORB was assessed in a cohort of reviews in which a core set of outcomes had been endorsed by trialists and the related Cochrane Review Group.

The aim of Chapter 2 was to present the methodology to assess suspected ORB in SRs and to estimate the prevalence of ORB in the Cochrane RA reviews by considering all eight outcomes in the COS.

In summary, from the results, firstly, all 21 reviews contained missing data on at least one of the eight core outcomes. Secondly, ORB was highly suspected in 275 (23%) of the 1205 evaluable outcomes from the 167 assessable trials scrutinised. Finally, ORB is a non-ignorable missing-data problem that could adversely affect the results of a SR.

In the presence of ORB, it is important to determine how robust MA conclusions are. If there is selective reporting and the presence of ORB is suspected, one possible solution is to contact the trialists in order to try and obtain the missing outcome data. However, in many cases this process is not feasible because missing data are unobtainable from trial authors, often because the authors are no longer easily contactable, the trials are old and data are not stored or readily available.

The impact of these missing data on the MA result should be examined through the use of a sensitivity analysis. In the statistical literature, there are three main statistical methods available for assessing the impact of ORB on MA results.

The first method has been termed maximum bias bound [70] and consists of calculating a pooled-effect estimate from the *n* studies reporting data and to calculate the bias-adjusted estimate by adding the value of this bound to the pooled-effect estimate. This approach assumes that larger studies (with small SEs) are more likely to be published than smaller studies (with larger SEs).

The second method, proposed by Copas et al. in 2014 [71], consists of checking if a paper in the area of interest does not report sufficiently well (or not at all) the particular outcome of interest using the ORBIT classification [28] method described in Chapter 2. This method explicitly models the ORB missing data mechanism according to the level of suspicion of ORB (again using the high/low classifications) from ORBIT. Fundamental to this approach is calculating the likelihood function [71] for each study being considered in the MA, taking into account that some of the outcome data are observed, some outcome data are missing due to a high ORB suspicion (i.e. the data are presumed missing because the result is not statistically significant) and some data are missing due to a reason associated with low suspicion of ORB, for example, the result has been measured as reported graphically but not necessarily analysed. The level of suspicion (high or low) also comes from the ORBIT classification [28] discussed in Chapter 2.

The third method is represented by MVMA, where the study effect estimates (which often relate to the treatment effect) for all available outcomes are jointly synthesised, while accounting for their w/s and between-study correlations [72]. In the presence of ORB, MVMA has been proposed as a possible statistical method to reduce the impact of ORB in a MA [72]. The main reason that the focus is on this MVMA method in this thesis is that MVMA allows for multiple outcomes to be synthesised simultaneously and the method accounts for the correlation across studies, which may add additional information to reduce ORB when there are missing outcomes. MVMA in this setting is largely unexplored and therefore sets the basis for this thesis.

The MVMA method utilises the w/s and (in the REMA setting) between-study correlations of the effects for the multiple outcomes in order to jointly synthesise outcomes from randomised trials. Furthermore, MVMA takes into account the correlation to gain more information than allowed in a traditional, UVMA of each outcome separately. From a practical sense, it is desirable to synthesise important outcomes in a review by performing one analysis, rather than analysing each outcome separately.

However, the lack of reporting of w/s correlations within trial reports often prevents the application of MVMA [73]. Various approaches for estimating the w/s correlation have been proposed in the literature. It has been demonstrated that, when IPD are not available, MVMA can be carried out by assuming a plausible value for each unknown correlation coefficient [74]. For example, there are some situations where the w/s correlation could be assumed as equal to zero [75]. These correlation values may also be derived using clinical considerations (e.g. elicitation of correlation information). In addition, the empirical correlation coefficient observed between treatment effect estimates across studies could be used.

If IPD are available for all studies, then the w/s correlations can be derived directly, which thus avoids reliance on study reporting [73]. Another simulation study has previously been undertaken in the area of MVMA for ORB. This study considered the UFMA and bivariate fixed-effects meta-analysis (BFMA) models and their estimation. The objective of the research was to demonstrate through a simulation study the estimation properties from BFMA when compared with those from two separate UFMA in situations with complete data and non-ignorable missing data according to an ORB mechanism [73]. The researchers obtained UFMA results that show ORB can substantially bias pooled estimates, and thereby over-estimate pooled treatment effects. They found that the results from the BFMA approach were encouraging and that the borrowing of strength (BoS), from taking into account the w/s correlation between outcomes running the model, can reduce the impact of ORB in a MA [72]. However, the researchers concluded that a major limitation of their study was that they generated the AD (effect estimates, and their w/s variances and correlations) directly, and so

did not generate IPD first. This has potential limitations. Firstly, the w/s correlations were always the same in each MA dataset, which in some situations may be unrealistic. Secondly, w/s variances were generated as known (i.e. they were not estimated from the IPD itself), which is also potentially unrealistic. Therefore, new simulations based on IPD are needed to verify the findings in a more realistic setting.

Undeniably, SRs and MA based on IPD are regarded as the gold standard [76] for evidence synthesis and have become increasingly common, having several advantages over meta-analyses based on AD. The advantages related to the use of the IPD in a MA involve all the aspects of the SR – from the trial inclusion, to the data quality, risk of bias assessment, analysis, and interpretation [77]. For example, the use of IPD would allow any unavailable w/s correlations to be calculated directly. Therefore, the use of IPD in MVMA should be considered.

The aim of this present study is to investigate the use of MVMA for reducing ORB in IPD MA, through simulation. This is the first IPD-based simulation study examining the impact of ORB reduction on pooled-effect estimates for meta-analyses.

For this simulation study, continuous *patient-level* responses are considered. IPD were generated for multiple outcomes in order to model the data across multiple studies using multivariate fixed-effects meta-analysis (MFMA) methods, for a number of different ORB scenarios, with comparison to UVMA.

The simulation study is motivated by the RA reviews that were previously considered in Chapter 2. Through the use of performance metrics, the objective is to demonstrate if MVMA offers an overall benefit over a standard UVMA approach when the w/s correlations and variances from each study can be computed directly.

In this chapter, the model specifications and estimation methods are presented for a UFMA model and a MFMA. The two-stage approach and models for fitting and analysing an IPD

meta-analytic approach are also described. In later sections, the simulation study structure, including the process for generating the data, incorporating missingness as a result of ORB, and the simulation performance indicators are both defined and described.

The chapter concludes with the results from the simulation study and discussion of the relative benefits of MFMA and any limitations.

3.2 Model specification and estimation

Before describing the MFMA model or considering IPD MA approaches, the standard UFMA and its estimation is firstly introduced. The UFMA model was also introduced in Chapter 1 in section 1.4.2. For UFMA model it is to be assumed that the effect estimates, their variances and correlations for a set of multiple outcomes are available from a set of randomised clinical studies to be meta-analysed. In general terms, there are to be *n* studies $(i = 1,....., n)$ and *m* outcomes $(k = 1,....., m)$.

3.2.1 Univariate fixed-effects meta-analysis

For a UVMA, there are two quantities of interest that are required from each study (*i*) for inclusion in the model: the treatment effect estimates and the w/s variance (or equivalently the SE).

Assume that for each study *i* there is a set of *m* treatment effect estimates $(y_{i1}, y_{i2},....., y_{im})$. Similarly, for each study *i* there is a set of *m* w/s variances for each outcome (say $s_{i1}^2, s_{i2}^2,....., s_{im}^2$).

Therefore the UFMA model assumes that the obtained estimates of the treatment effect for the i^{th} study are normally distributed about a common, fixed, true effect and variance, which are assumed to be known.

Of prime interest is estimating this fixed true effect value. For (*m*) separate outcomes, there are (*m*) separate normal distributions, as Equation 3.1 demonstrates [17]:

$$\begin{aligned} y_{i1} &\sim \mathrm{N}\left(\mu_1, s_{i1}^2\right) \\ y_{i2} &\sim N\left(\mu_2, s_{i2}^2\right) \\ y_{im} &\sim N\left(\mu_m, s_{im}^2\right) \end{aligned} \qquad (3.1).$$

Each model parameter (for each outcome separately) is typically estimated using ML estimation. The final (pooled) estimate of the MA is a weighted average of the estimates of the individual studies. In UFMA, model weights are a function of the variability of each study; in other words, the inverse of the variance. For example, for outcome one, the formula for the overall pooled estimate for outcome one is the following [17]:

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n} \frac{y_{i1}}{s_{i1}^2}}{\sum_{i=1}^{n} \frac{1}{s_{i1}^2}} \qquad (3.2).$$

3.2.2 Multivariate fixed-effects meta-analysis

For MFMA, all outcomes are analysed simultaneously [76]. In addition to the two quantities required for a UVMA from each study (treatment effect estimate and the w/s variance), the w/s correlations are also required. The w/s correlations indicate the strength of the association between the outcome estimates within a study, and these are assumed to be known. These w/s correlations are rarely reported in primary studies, but, in Section 3.3, a process for deriving these quantities from each study is described when IPD are available. Therefore, the general form of the MFMA model follows the multivariate normal distribution and is shown in Equation 3.3 [76]:

$$\mathbf{Y}_i \sim N(\boldsymbol{\mu}, \mathbf{S}_i) \qquad (3.3)$$

In Equation 3.3 $\mathbf{Y}_i$ represents the vector for the treatment effect estimates, while $\boldsymbol{\mu}$ is the vector of the fixed true treatment effects for the outcomes and $\mathbf{S}_i$ represents the w/s variance-covariance matrix of the effects' estimates. For *m* outcomes, the model in Equation 3.3 is equivalent to:

$$\begin{pmatrix} y_{i1} \\ y_{i1} \\ \vdots \\ y_{im} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_m \end{pmatrix}, \mathbf{S_i} \right) \qquad \mathbf{S_i} = \begin{pmatrix} s_{i1}^2 & s_{i12} & \cdots & s_{i1m} \\ . & s_{i2}^2 & \cdots & s_{i2m} \\ \vdots & \vdots & \ddots & \vdots \\ . & . & \cdots & s_{im}^2 \end{pmatrix} \qquad (3.4).$$

The entries of the matrix $\mathbf{S}_i$ on the main diagonal are the w/s variances $(s_{i1}^2, s_{i2}^2, ..., s_{im}^2)$, which are assumed to be known. The entries outside the main diagonal are the w/s covariances, which are also assumed to be known.

ML estimation method can again be used to estimate the vector of pooled estimates $\hat{\mu}$, for example, using the STATA MVMA module 'MVMETA' [78]. One of the benefits of the model in Equation 3.3 is that it can accommodate studies that do not report on data for all outcomes by utilising the w/s correlation. In the field of MVMA, the impact the w/s correlation has on the estimation of the treatment effects is often termed borrowing of strength. BoS can typically depend on the strength of the w/s correlations and also the size of the w/s variances between outcomes [76]. In the presence of missing study data, BoS can be particularly large, as this is equivalent to arbitrarily setting a w/s variance of infinity for missing outcomes in such studies. The concept of 'borrowing strength' is returned to in Section 3.9, where it is considered as an important performance indicator in the simulation study.

3.3 Individual participant data multivariate meta-analysis

The motivating RA example underpinning this thesis considers only multiple continuous outcomes and hence this thesis considers only facilitating both the univariate and multivariate models (Equations 3.1 and 3.3 respectively) for this particular scenario. Consequently, the IPD will be used to obtain the effect estimates, their within study variances and covariances from each study under the assumption that all outcomes are continuous. Focus is also given to the two-stage IPD-MA approach suggested by Riley et al. 2010 [76], although a one-stage approach is also possible.

Under a two-stage approach, when IPD are available from *n* studies, in the first stage, each of these studies is analysed separately to obtain the AD from each study, and then, in the second stage, the models in Equation 3.1 and Equation 3.2 can be applied.

3.3.1 First-stage

As with many RA trials, most are two-arm, parallel group trials comparing a treatment (T) against a comparator drug (C). The model specification for the first stage of an IPD-MA is dependent on the specific metric of analysis. The following considers the final score analysis that will be used in the simulation, but other specifications are also available, for example, change from baseline [79].

At the end of each study, for a final score analysis, the j^{th} patient in the i^{th} study will provide a final score for each outcome. Each individual patient response will be referred to as $\mathbf{z}_{ijk}$, i.e. the response value for outcome (*k*) from patient (*j*) in the i^{th} study. For each outcome and each trial separately, the general form of the model for the final score analysis becomes:

$$z_{ijk} = \alpha_{ik} + \beta_{ik} T_{ij} + \varepsilon_{ik} \quad \text{where} \quad \varepsilon_{ik} \sim \mathrm{N}(0, \sigma_{ik}^2) \quad (3.5)$$
$$\text{and} \quad \mathrm{cov}(\varepsilon_{i(m-1)}, \varepsilon_{im}) = \sigma_{(m-1),m}$$

In Equation 3.5, α_{ik} is the control group effect (intercept) for outcome *k*, β_{ik} is the treatment effect for outcome (*k*) and study (*i*), while T_{ij} identifies the treatment group indicator for the *i*th study and the *j*th patient. In this joint model (Equation 3.5), the residual errors are equally distributed as a Normal with mean 0 and variance σ_{ik}^2, which represents the residual variance of outcome *k* in trial *i*, after accounting for the treatment effect. The residual errors are also correlated and the covariance $\mathrm{cov}(\varepsilon_{i(m-1)}, \varepsilon_{im}) = \sigma_{(m-1),m}$ is different from 0 and equal to $\sigma_{(m-1),m}$.

In the simulation study, just three continuous outcomes will be considered (*k*=3), where *k*=1 for the outcome TJC, *k*=2 for the outcome SJC and *k*=3 for pain, as an example from the RA core set of outcomes defined in Chapter 2. Consideration of only three outcomes is for simplicity. For the motivating example, the model in Equation 3.5 can be fitted as a trivariate model with a joint model specification, where each patient contributes to three follow-up responses (one for each outcome *k*), to a single model containing all the three outcome jointly [79]. Therefore, the joint model fitted in each trial – taking into account all three outcomes as in this example – assumes the following structure:

$$\begin{pmatrix} z_{ij1} \\ z_{ij2} \\ z_{ij3} \end{pmatrix} = \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \alpha_{i3} \end{pmatrix} + \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \beta_{13} \end{pmatrix} \times T_{ij} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix}$$

where (3.6)

$$\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_i \right) \qquad \Sigma_i = \begin{pmatrix} \sigma_{i1}^2 & \sigma_{i12} & \sigma_{i13} \\ \sigma_{i21} & \sigma_{i2}^2 & \sigma_{i23} \\ \sigma_{i31} & \sigma_{i32} & \sigma_{i3}^2 \end{pmatrix}$$

Of interest from the model shown in Equation 3.6 are the estimation of the treatment effects vector $\boldsymbol{\beta}_i$, the associated SEs of $\boldsymbol{\beta}_i$ and the patient's level (residual) variance-covariance parameters $\boldsymbol{\Sigma}_i$. At the second stage of the IPD model, the SEs will be used to calculate the variances to run the MFMA model in Equation 3.2. It should be noted that, for the future implementation of UFMA or MFMA models, the vector of $\boldsymbol{\beta}_i$ estimates in Equation 3.6 contains exactly the same values contained in the vector of the treatment effect estimates $\mathbf{Y}_i$ of Equation 3.2. In order to estimate these quantities, the model in Equation 3.3 (and Equation 3.4) can be fitted using SAS 'proc mixed' using REML as the method of estimation.

At this stage of the IPD-MA it is fundamental to recall that, having calculated the covariances, for example, for outcome one and outcome two, σ_{i12}, between the patient outcome responses, and the SEs σ_{i1} and σ_{i2}, it is also possible to calculate the w/s correlations ρ_{wi12} between individual outcome responses [80], as follows:

$$\rho_{wi12} = \frac{\sigma_{i12}}{\sigma_{i1}\sigma_{i2}} \qquad (3.7)\,.$$

3.3.2 Second-stage

From stage one, the treatment effect estimates for each outcome *k* in each study *i*, and their w/s variance-covariance matrices (containing their variances and w/s correlations, see Equation 3.7) are obtained. Having estimated these parameters it is possible to run the UFMA model (Equation 3.1 and Equation 3.2) [79] and therefore obtain the pooled-effect estimates, their SEs and the 95% confidence intervals. It is also possible to apply and fit the MFMA model (Equation 3.3 and Equation 3.4) [79] and again obtain the pooled-effect estimates, their SEs and the 95% confidence intervals.

3.4 Description of the simulation study

As previously mentioned, in this simulation study, three continuous correlated outcomes from the COS [25] endorsed by the RA SRs are considered (outcome 1: TJC, outcome 2: SJC and outcome 3: pain). TJC and SJC are discrete numerical variables (treated and analysed as continuous outcomes), while pain is measured on a continuous scale. For all these outcomes, the treatment effect is defined as the MD between the groups of patients randomised to control group treatment (C) and the group of patients receiving active treatment (T). In order to introduce correlated patient data there is also a need to define true *patient-level* correlations (ρ_p) between the outcomes. This simulation study considers a variety of different *patient-level* correlations between outcomes, including a set of correlations for the outcomes considered from a real IPD dataset of RA. Differing numbers of studies in the MA may also have an important factor on results (in this study we consider five and ten studies, as these reflect the range of studies found in reviews of RA) as well as different severities of missingness according to known ORB mechanisms. The simulation study examines 25 different scenarios in total, categorised (i) to (xxv) (Table 3.1). The details on how the data were generated and missing data introduced are provided in Sections 3.5 and 3.7 respectively.

Table 3.1 Scenario used in the simulation

Scenario	Assumed *patient-level* correlation: ρ_p	Description
Complete		
i	0	No missing data
ii	0.2	
iii	0.5	
iv	0.8	
v	RA correlations from a real clinical trial (RA_corr)	
ORB		
vi	0	Set missing all the *study results* showing a benefit to alternative treatment (i.e. study outcome with a positive MD)
vii	0.2	
viii	0.5	
ix	0.8	
x	RA correlations from a real clinical trial (RA_corr)	
xi	0	Set missing 15%, 20% and 30% respectively for outcomes Y_1, Y_2 and Y_3 of the *study results* non-significant after 'first stage' IPD-MA
xii	0.2	
xiii	0.5	
xiv	0.8	
xv	RA correlations from a real clinical trial (RA_corr)	
xvi	0	Set missing all the *study results* that are non-significant after the 'first stage' IPD-MA
xvii	0.2	
xviii	0.5	
xix	0.8	
xx	RA correlations from a real clinical trial (RA_corr)	
xxi	0	Set missing all the *study results* that are either non-significant or showing a benefit to alternative treatment
xxii	0.2	
xxiii	0.5	
xxiv	0.8	
xxv	RA correlations from a real clinical trial (RA_corr)	

RA_corr: 0.67 (Y1 & Y2), 0.55 (Y1 & Y3), 0.38 (Y2 & Y3)

3.5 Generating the simulated data

The purpose of this section is to explain how the IPD were generated for each of the 1000 MA simulation datasets *nsim* ($nsim = 1,.....,1000$).

The true effect estimates (MDs) were set equal to -5 for outcome one (TJC), -3 for outcome two (SJC) and -6 for outcome three (pain). While the choice of treatment effect estimates

maybe seen as an arbitrary one in a simulation study, typically these were set such that they were the 'average' treatment effect values observed in the RA reviews. As an example, for TJCs, an interpretation of a MD of -5 would reflect that 'on average' five fewer TJCs were observed in the treatment group than in the control group.

The number of studies (*n*) within each MA and each simulation scenario was fixed at five studies and ten studies. In five-study scenarios, the sample sizes (*N*) for each study ranged from 30 to 70 in each treatment arm, and this was from 10 to 100 in each treatment arm for meta-analyses containing ten studies. Sample sizes were fixed for each study across all simulations. For scenarios where *n* = 5 studies were simulated, the first study contained 30 patients in each treatment arm, increasing in increments of 10 additional patients up to study 5, which had 70 patients. When considering scenarios where *n* = 10 studies were simulated, the first trial had 10 patients in each treatment arm, and then went up in increments of 10, where study 10 had 100 patients in each arm. The ratio between the sample sizes of patients in each treatment arm was set as 1:1.

Based on the number of studies and sample sizes, two sets of data were generated for each outcome, one for the group receiving treatment (T), and the patients receiving placebo in the control group (C). The response variable for each outcome was then simulated from a multivariate normal distribution and a vector of MDs (representing the final score) for the three outcomes generated. The MDs between the group of patients treated and the control group of patients were fixed as described above.

The multivariate normal distribution through the specification of a variance-covariance matrix structure allowed the IPD for the three outcomes generated for each study to be correlated within each simulation. The *patient-level* correlations ρ_{pi} were fixed across all studies and simulations and set to be 0, 0.2, 0.5, and 0.8 between each pair of the three outcomes. An additional *patient-level* correlation scenario was also considered (RA_Corr). This resembled the actual *patient-level* correlation between the three outcomes from a real RA IPD dataset.

These *patient-level* correlations were as follows: (TJC/SJC =0.67; TJC/pain=0.55; SJC/pain=0.38).

Furthermore, MVMA needs the variance-covariance matrix to be defined. In this simulation, different size of variances were considered for each study, where study one, with the lower sample size, has been allocated the lower variances for each outcome and study five, with the higher sample size, has been allocated the higher variances for each outcome considered.

Considering the variance-covariance matrix, the idea was to generate IPD as similar as possible to the actual data observed during the analysis of the SRs. First of all, as in the analysed SRs, where it had been noticed that each outcome was characterised by different variability, three different variances were considered for each outcome of this simulation study. Secondly, in analysing the SRs it had been observed that the treated group of patients and the control group of patients were presenting different variability. Therefore, considering each single outcome, the IPD were simulated taking into account this difference, setting two different sets of variances for treated and control group. As shown in the Appendix B, for studies with higher sample sizes, higher variances were considered for the groups of patients considered in the simulated randomised trials.

The SAS code to generate the DATA is provided in Appendix B. A fuller specification of the parameters used including the covariance parameters is included.

3.6 First stage of individual participant data meta-analysis

After simulating the IPD for each study and each simulation, of interest is the calculation of the estimates associated with the treatment groups. Furthermore, the attention needs to be paid to the estimate of the MD, across the groups of patients randomised to treatment or control group, for the outcomes considered. Also needed are the estimated w/s variances and w/s correlations defined in Equation 3.5 under the w/s variance-covariance matrix. At

this stage, the aim is to apply the joint fixed model (Equation 3.6 or Equation 3.7 as in the motivation example of this work of simulation there are three outcomes considered) as described in Section 3.3.1, through the introduction of dummy indicator variables, which identify which patient belongs to which group.

3.7 Generating the missing outcome data

Following on from Section 3.5, at this stage, the IPD data generated are complete. Missing data are generated according to the varying degrees of ORB severity mechanisms outlined in Table 3.1. The mechanism in scenarios xi-xv resembles the proportion of missing study data found in previously analysed RA reviews for the outcomes considered. Scenarios xxi to xxv are the most extreme cases of ORB missingness.

Once the estimates of the effects (MD) for each study and each simulation and SEs and p-values associated with them (from stage 1 of the IPD-MA) had been obtained, missing data at the study level were introduced, according to the various scenarios. In other words, introducing missing data at the study level means that some of the results of the studies included in the MA that have been simulated are omitted. The objective was to replicate, as far as possible, realistic scenarios of ORB. Therefore, missing study data in the 1000 simulated datasets were simulated.

For example, for scenarios vi to x (Table 3.1), as it has been defined, the objective was to introduce missingness in correspondence to all the study results showing a benefit to alternative treatment (i.e. those with a positive MD). To introduce this type of missingness, following the results from the first stage of the IPD-MA, all the study-level results showing a positive MD for all 1000 simulated datasets were omitted. Study data for all other scenarios were omitted according to the mechanisms described in Table 3.1.

3.8 Fitting the meta-analysis model

The first stage therefore provides treatment effects' estimates and their variances, for each outcome in each study, and their w/s correlations. After the introduction of missing data at the study level (except in the case of the complete case analysis), it is then possible to implement the UFMA model following Equation 3.1 or MFMA using Equation 3.3. UFMA and MFMA models will give the pooled-effect estimates and their SEs for each outcome considered for each of the 1000 simulated MA datasets.

For computational convenience, some studies were incorporated into the MA by allocating them an arbitrary value (e.g. set the treatment effect to zero) with a very large within-variance (e.g. 1,000,000) and w/s correlation equal to zero. This technique is called data augmentation and replaces the missing outcome value with values that have negligible weight and information during estimation.

Confidence intervals and hypothesis tests immediately followed using the conventional procedure by assuming that the estimates $\hat{\mu}_k$ obtained from the UFMA model in Equation 3.1 or the MFMA model in Equation 3.2 are normally distributed.

3.9 Assessment of performance

The performance of the estimates was assessed in terms of bias, mean SE, mean square error (MSE), coverage, power and BoS.

Considering $(nsim = 1,.....,1000)$ the number of simulations and $(k = 1,.....,3)$ the number of outcomes, the bias for outcome *k* is defined as [17]:

$$Bias\left(\hat{\mu}_k\right) = \frac{\sum_{nsim=1}^{1000}\left(\hat{\mu}_{knsim} - \mu_k\right)}{1000} \qquad (3.9)$$

Therefore, the bias is calculated as the average of the differences between the estimator expected value $\hat{\mu}_{knsim}$ for each simulation and each outcome and the true value of the parameter μ_k being estimated.

Furthermore, the average standard error (SE_{sim}) and mean squared error (MSE_{sim}) were calculated to assess precision, taking the average of all the SEs and all the mean squared errors associated with the estimates obtained for each simulation with reference to the outcomes analysed.

Then the coverage and the power associated with the overall estimates were calculated.

The coverage was calculated as the percentage of simulated datasets for which the 95% confidence interval for an outcome's treatment effect estimate contained the true effect estimate μ_k .

Power was calculated as the percentage of simulated datasets for which the 95% confidence interval for an outcome's treatment effect estimate did not contain zero [17].

For all scenarios, the BoS was also calculated with the objective to compare the results obtained from UFMA and MFMA to see how much influence the utilisation of the w/s correlation has on the BoS at the MVMA level. For this, $\overline{\mathrm{Var}_{\mathrm{MFMA}}\left(\mu_k\right)}$ needed to be defined as the average of the estimated variances for outcome *k* for the MFMA model and $\overline{\mathrm{Var}_{\mathrm{UFMA}}}\left(\mu_k\right)$ as the average of the estimated variances for outcome *k* from the UFMA model. The formula that we used to calculate BoS is reported in Jackson et al. [80]. In the following formula (Equation 3.12), the BoS will be calculated as a percentage, as follows:

$$BoS\left(\hat{\mu}_k\right) = \left(1 - \frac{\overline{\mathrm{Var}_{\mathrm{MFMA}}\left(\mu_k\right)}}{\overline{\mathrm{Var}_{\mathrm{UFMA}}\left(\mu_k\right)}}\right) * 100 \qquad (3.10)\,.$$

It is know from the literature that large w/s correlations allow BoS across outcomes, and this produces slightly narrower 'confidence intervals' as a result of smaller SEs [18].

3.10 Simulation Results

The aim is to study the impact of bias in the pooled treatment effect estimates for multiple outcomes using UFMA and MFMA, and to determine whether the BoS from MVMA can reduce the impact of ORB.

The results of the simulation studies are provided in a separate appendix (Appendix C). Table C.1 in this appendix provides the simulation results for scenarios with no missing data. The results for scenarios with missing data at the study level, as a result of ORB, are shown in Tables C.2 to C.5 of this appendix.

The purpose is now to discuss the new findings from this present simulation study, in particular focusing on the utilisation of w/s correlation. The trend in the parameters considered for the evaluation of the results will be described, and in particular the variation and the trend in the bias of the estimates, SEs (SE_{sim}), coverage, mean squared error (MSE_{sim}) and power will be examined. The variation of percentage of BoS for all outcomes will also be presented and discussed.

Data augmentation was needed for a number of the simulated MA datasets; the number of datasets this method was applied to, increased, as the amount of missingness became more severe (Tables C.1-C.5 in Appendix C). In the tables of the appendix C, μ_1 refers to the outcome 1 (TJC), μ_2 refers to the outcome 2 (SJC) and finally μ_3 refers to the outcome 3 (pain).

3.10.1 Comparison between univariate fixed-effects meta-analysis and multivariate fixed-effects meta-analysis

For all scenarios, complete cases and missing ORB mechanisms, there were no discernible differences in power; hence, it was difficult to make a judgement on the benefit of MFMA using this performance criterion. Notably, the level of power achieved was high, and in many cases 100% power was achieved. The high levels of power observed are perhaps the consequence of performing IPD-MA with true effect sizes far from zero.

For complete-case data, the estimates for all outcomes were relatively unbiased for both UFMA and MFMA (Table C.1 Appendix C). As the missing ORB data mechanism became more severe, the estimates became more biased. The negative biases observed in this instance were indicative of the pooled treatment effect overestimating the true treatment effect. The most severe biases occurred when both non-significant results and results showing benefit to the alternative treatment were excluded (Table C.5 Appendix C). There was a tendency for the biases to be greater for the estimates obtained in scenarios where there were ten studies compared to those with five studies. It is expected that this is a result of the additive effect of excluding more studies from the meta-analyses, resulting in more biased pooled estimates. Similarly, there was a tendency for outcome three to be more biased than outcome two and outcome two to be more biased than outcome one, a result of differing levels of missingness in each outcome. In terms of bias reduction, there was a benefit from using MFMA over UFMA in all scenarios where at least all non-significant study results were removed from the MA (Tables C.4 and C.5 in Appendix C). The benefit seemed to be greater as the *patient-level* correlation between outcomes increased and marginally improved when there were fewer studies in the MA (five studies as opposed to ten). As an example, from Table C.5 Appendix C, where five studies were simulated for each simulation, with a *patient-level* correlation of 0.8 between outcomes, the bias reduction in outcome two was about 26% when comparing the UFMA estimate (-1.129) with the MFMA estimate (-0.838). When a lower *patient-level* correlation value of 0.2 between outcomes was used, the

bias reduction was about 2% when comparing the UFMA result (-1.160) with the MFMA result (-1.140).

When the missing ORB data mechanism was less severe (Tables C.2 and C.3 in Appendix C), the MFMA still outperformed the UFMA in the majority of cases. There were some instances when the UFMA was slightly improved, but this nearly always occurred in outcome one where the bias was generally smaller due to fewer studies being missing from the analysis. The benefit gain in the UFMA was also only negligible.

In terms of standard errors (SE_{sim}) and MSE_{sim}, the pooled estimates were always more precise using MFMA, resulting in smaller standard errors and mean square errors, with larger gains when there were missing data as a result of ORB and when the patient level correlation was increased. As an example, considering outcome two, for ten studies with a *patient-level* correlation of 0.8 (Table C.5 in Appendix C), the average standard error for the UFMA approach was 1.178 compared with 0.950 for MFMA, an increase in precision of 19%. These findings are consistent with those found in previous simulation studies, albeit these considered BFMA.

This pattern in the increase in precision is also reflected in the BoS performance indicator that was not previously considered in other simulation studies. The more severe the ORB missing data mechanism and the stronger the *patient-level* correlation, the bigger the gain in BoS when applying the MFMA method. For example, considering the most severe ORB scenario (Table C.5 in Appendix C), where the *patient-level* correlation is set to 0.8 and the number of studies at ten, the %BoS is 8.6%, 31.7% and 19.3% for outcomes one through to three respectively. Considering the same level of studies and same *patient-level* correlation for a less severe ORB mechanism, the %BoS is much lower, for example, 4.3%, 7.6% and 5.6% for the three outcomes respectively (Table C.5 in Appendix C). Considering the impact on the *patient-level* correlation comparing the most extreme ORB mechanism, again with ten studies but with a *patient-level* correlation set to 0.2, the %BoS is 1.7%, 3.3% and 2.3%

(Table C.5 in Appendix C) for the three outcomes, far lower than the 8.6%, 31.7% and 19.3% for a *patient-level* correlation of 0.8.

For complete-case scenarios, coverages were hovering around the nominal 95% for both UFMA and MFMA approaches. Similarly, nominal coverages were maintained for both methods when the ORB mechanism was less severe (benefit of alternative treatment removed, Table C.2 in Appendix C). There was a marked drop in coverage as the missing ORB mechanism became more severe; this was particularly noticeable for ten studies compared to five. For these situations, there were quite a few instances where the UFMA marginally outperformed the MFMA approach. For example, taking into account the scenario where 15/20/30% of study results were suppressed from TJC/SJC/Pain respectively if they were non-significant and considering the situation where n = 5 studies, the UFMA coverage results were closer to nominal (93.3% for outcome 1, 94.7% for outcome 2 and 92.1% for outcome 3) than the MFMA coverage results (93.1% for outcome 1, 94.3% for outcome 2 and 84.0% for outcome 3). Nevertheless, in scenarios where the ORB mechanism was most severe, MFMA was preferable, particularly when the *patient-level* correlation increased. For example, taking into account the scenario where all studies that have non-significant results and show benefit of alternative treatment are excluded and considering the situation where n = 5 studies, the UFMA coverage results were lower (91.7% for outcome 1, 85.7% for outcome 2 and 89.8% for outcome 3) than the MFMA coverage results (91.9% for outcome 1, 88.8% for outcome 2 and 90.6% for outcome 3).

3.11 Discussion

3.11.1 Research problem

The objective of this present study was to investigate the use of MVMA for reducing ORB in IPD-MA, through simulation. This is the first IPD-based simulation study examining the impact of ORB reduction on pooled-effect estimates for meta-analyses.

3.11.2 Rationale for individual participant data meta-analysis

In this chapter, both the UFMA and MFMA approach for meta-analysing IPD were considered. The MVMA approach offers a novel way of jointly synthesising multiple outcomes across multiple studies. The added benefit of having access to the IPD is that the often-unreported w/s correlations could be computed directly from the data [79].

The IPD-MA simulation study that was performed demonstrated that the MFMA was preferential over UFMA across the majority of performance criteria and scenarios considered. The benefit was particularly noticeable as the ORB missing data mechanism became more severe and the *patient-level* correlation between the outcomes increased. This study also utilised the BoS performance measure, which has not been previously considered in related simulation studies. This enabled the quantification of the relative benefit in key performance criteria when comparing UFMA with MFMA.

In this simulation study, a two-stage IPD-MA was applied for convenience and simplicity. This approach means that the estimates of the treatment effects, variances and w/s correlation at the first stage for each separate study can be first estimated for each separate simulation, allowing these estimates to be used in the UFMA or MFMA models at the second stage of the IPD-MA. From a statistical perspective, this approach avoided possible problems of convergence and is particularly feasible when data are missing in some outcomes [79]. Nevertheless, a one-stage MA approach could have been considered, analysing all IPD together from all studies to perform a single IPD analysis with a mixed-effects model [79].

In this thesis study, continuous outcomes were simulated. Other studies have been considered and analysed with multivariate IPD-MA; furthermore, other outcomes such as binary, survival and mixed have also been considered.

3.11.3 Limitations of the study

The first limitation of this study is that we only considered FEMA approach. In many ORB scenarios, this may not be a realistic assumption in practice because ORB may hide the true extent of any underlying heterogeneity. This may suggest that a REMA approach ought to be the default methodology in the presence of severe ORB.

A second limitation of this simulation study is that the number of studies (*n*) within each MA and each simulation scenario was fixed across the five studies and ten studies. This could have had an effect on the results, reducing the variability of the response variable between studies across all the 1000 simulations considered. Taking this into consideration, some of the differences between five and ten study scenarios may not be directly comparable.

A third limitation of this simulation study is that, when UFMA was applied at the *second* stage of IPD-MA, the known treatment effect estimates that are required to calculate the UFMA model are the estimates of the MD calculated from the IPD model implemented at the *first* stage. The same estimates of the MD and the variances were used also to implement the MFMA at the second stage.

Chapter 4 will investigate an alternative method to analyse the IPD at the first stage that could make the results between UFMA and MFMA more different in terms of performance criteria such as Bias or Standard Error.

3.11.4 Conclusion

The main conclusion to be drawn from the simulation study that was performed is that the results are encouraging and the use of MFMA over standard UFMA is strongly recommended. There is also an argument to suggest that the method is no more complex to understand than a standard IPD UVMA, if this is being considered. Nevertheless, in the

presence of severe ORB mechanisms, it should be noted that, while MFMA can reduce the bias compared to standard UFMA, the results can still remain largely biased.

3.11.5 Future applications of univariate fixed-effects meta-analysis and multivariate fixed-effects meta-analysis to missing data scenarios simulated at individual participant data level

In Chapter 4, the purpose is to retrace the simulation study by applying the UFMA and MFMA methods to scenarios where missing data have been introduced at the level of IPD, according to various types of missingness: missing completely at random (MCAR) and missing at random (MAR).

Chapter 4 - Multivariate fixed-effects meta-analysis to adjust for missing data in individual participant data: a simulation study

4.1 Background

Chapter 3 considered both the UFMA and MFMA approaches for meta-analysing individual IPD. The chapter investigated the use of MVMA for reducing outcome reporting bias (ORB) in IPD-MA. A series of ORB scenarios was considered. In particular, an IPD-MA simulation study proved that the MFMA was preferable to UFMA across the majority of performance criteria and scenarios considered. Furthermore, the previous chapter also showed that the results obtained from the application of UFMA and MFMA were encouraging. The conclusion was that the use of MFMA would be recommended. However, in the presence of severe ORB mechanisms, it has been noted that, while MFMA can reduce the bias compared to standard UFMA, the results can still remain largely biased. Ignoring missing data will lead to bias if the missing data mechanism is related to both the treatment and the unobserved outcome (e.g. missing values are more likely in one treatment arm because it is not effective). Missing data is a common issue in randomised controlled trials and threatens a trial's ability to yield definitive conclusions [81]. Missing outcome data in trials can be due to many reasons and the explanations for it may follow the three basic missing data mechanisms: MCAR, MAR and missing not at random (MNAR) [82].

In the presence of missing outcome data, trial authors often ignore the missing data and perform a complete case analysis [83]. In the CONSORT statement, point 13b states "for each group, losses and exclusions after randomisation, together with reasons" should be included within the trial report [83]. This is not adhered to in a large number of trials within the study [83]. It is difficult to suggest a gold standard for missing data handling, as the appropriateness of a method is dependent on the unique nature of missing data within each

individual trial [84]. However, by carrying out a complete case analysis or eliminating certain patients based on level of missingness or prognostic factor we are making a bold assumption that the data we exclude is missing completely at random [84]. This is rarely, if ever, the case in practice [84]. For example, in a recent review of 100 clinical trials with a longitudinal outcome, nearly a third of the trials (32/100) performed only a complete case analysis despite there being missing outcome data [83]. Of those that did apply an imputation method, only half were deemed to have used an adequate method, in particular a simple imputation called LCOF/FOCB/Baseline Carried Forward [83]. Here, in Chapter 4, this work is extended by simulating a different mechanism for dealing with missing data at the patient level and applying UFMA and MFMA.

MVMA has been considered and applied in both this chapter and in Chapter 3. The importance of its application is due to the following reasons. First of all, MVMA utilises the within-study correlations of the effects for the multiple outcomes in order to jointly synthesise outcomes from randomised trials. Furthermore, MVMA takes into account the correlation to gain more information than allowed in a traditional, UVMA of each outcome separately. In a practical sense, it is desirable to synthesise important outcomes in a review by performing one analysis, rather than analysing each outcome separately [17].

The concept of IPD was introduced in the previous chapter and relates to the data recorded for each subject in a study. A set of IPD from multiple studies often comprises a large sample size of patients. The notion of IPD is in contrast to the term AD, which relates to information averaged or estimated across all individuals in a trial. It is important to bear in mind that AD are derived from IPD themselves, so IPD can be considered the original source material. The objective of IPD-MA is to summarise the evidence on a particular clinical question from multiple related studies, such as whether a treatment is effective [76]. IPD-MA allows systematic reviewers to obtain the entire within-study variance/covariance matrix in each study, alleviating the reliance on this information being reported in the trial manuscripts. IPD may be not available for some studies. In this situation, one solution is to use the within-study correlations derived from IPD studies to give a likely value of the within-study correlation in AD studies [17]. It has been discussed in the literature that one of the

potential advantages of using IPD for a MA is that missing data can be observed and accounted for at the individual level by applying imputation methods. Examples include using simplistic methods such as single imputation, for example, by using mean substitution, or more sophisticated methods such as the application of the expectation-maximisation (EM) [82] algorithm or multiple imputation [82], prior to synthesising the data. In this way the sensitivity of the pooled effects estimates from the imputed data sets can be assessed against the available case or complete case data. When only AD is available, systematic reviewers have little control over how to adjust for the missing data in their meta-analyses, other than to make an assessment of the potential impact of the missing data when assessing the risk of bias in the study. This also assumes that the amount of missing data is known from the primary study report, which may not always be the case, especially at the individual outcome level.

The simulation study presented in Chapter 3 does not take into account missing data at the patient level, but it does consider it at the study level. With the existence of IPD, it is possible to deal with missing data at the patient level. Therefore, the aim of this current chapter is to simulate IPD for each study and within each simulated dataset, to introduce at the patient level different mechanisms for missing data, and finally to apply IPD-MA to the simulated data set. As in Chapter 3, 4 UFMA and MFMA models will be applied. The purpose of this chapter is to demonstrate is that MVMA could improve the estimates in the presence of missing data at the participant level.

4.2 Steps of the individual participant data simulation analysis applied to missing completely at random and missing at random scenarios

The aim of this section is to describe the present simulation study. The objective of this study is to simulate different missing data mechanisms at the IPD level and to analyse these missing data with the IPD models that will be described in section 4.3. All the characteristics in this simulation study remain the same as those described in Chapter 3; the only difference introduced in this chapter is that the missing data mechanism are considered at the

individual participant level rather than the study level. These are the steps that have been followed to obtain the results that will be presented and discussed in this chapter:

- *Step 1*: IPD simulation. The first step of IPD-MA must estimate the treatment effects for each outcome in each trial. Therefore, IPD for three correlated outcomes (outcome 1 was TJC, outcome 2 was SJC count and outcome 3 was pain) were simulated following a multivariate normal distribution to simulate continuous outcomes. The motivating example for this IPD simulation study was the assessment of a Cochrane Systematic Review for the treatment of Rheumatoid Arthritis previously analysed. Then different mechanisms – a MCAR mechanism and a MAR mechanism – in IPD were simulated and studied.

- *Step 2*: IPD-MA. The interest was in calculating the estimates associated with the treatment effects, measured as a mean difference across the groups of patients randomised to treatment or control group, for the three outcomes considered. To apply IPD-MA, a two-stage IPD-MA approach was chosen [79].

  i. During the *first stage* of IPD-MA, it was considered using two different approaches. The first method consisted of applying a *separate* model, which is described in section 4.3.1.1 of this chapter. The second approach consisted of implementing a *joint* model, which is described in section 4.3.1.2 of this chapter. The first stage therefore provides treatment effect estimates and their variances, for each outcome in each study, and their within-study correlations. Treatment effects and their variances are estimated from SAS PROC MIXED [85] with the REML method (Appendix D).

  ii. The first stage therefore provides treatment effect estimates and their variances for each outcome in each trial, and their within-study correlations. Then, at the second stage of the IPD it was possible to implement the UFMA model described in section 4.3.2.1 of this chapter or MFMA, which will be described in

section 4.3.2.2. UFMA and MFMA models will give the pooled-effect estimates and their standard errors for each outcome considered for each of the 1000 simulated MA datasets. In applying the UFMA to obtain the pooled mean difference estimate at this stage, the results from the two models implemented at the first stage have been taken into consideration. The first model was called the separate model (UFMAs) and the second model was called the joint model (UFMAj). To obtain the desired results, the mvmeta command in STATA [78] was utilised. See Appendix D for more details about the STATA code.

- *Step 3*: the results obtained from the UFMA and MFMA have been summarised with performance indicators, as will be described in section 4.7 of this chapter. The results are contained in Appendix E and reported in section 4.8 of this chapter and discussed in section 4.9.

4.3 Defining individual participant data meta-analysis

In Chapter 3, the two-stage IPD-MA was defined in detail. As defined in the previous chapter, the most frequent scenario in many rheumatoid arthritis trials is that most of them are two-arm parallel group trials comparing a treatment (T) against a comparator drug or control (C). The model specification for the first stage of an IPD-MA is dependent on the specific metric of analysis. Again, the final score analysis to be used in this simulation is considered. The setup of the simulation study is similar to that in Chapter 3, with the difference being only in introducing missingness at the participant level rather than the study level; an additional modelling approach is also considered here for comparison.

In Chapter 3, when UFMA was applied, the *joint* model was considered but in this chapter a *separate* model is also investigated. Here there is the need to introduce the difference between the joint model and the *separate* model because each outcome is investigated separately within each study by fitting n different UFMAs, where *n* is equal to the number of the outcomes considered in the simulation study.

In situations with a large proportion of incomplete outcome data, as in the scenarios that will be considered in this chapter, it is reasonable to assume that applying a *separate* model instead of applying a joint model could lead to different treatment effect estimates and variances. In the next sections, the focus will be on the description of the differences between these approaches.

4.3.1 First stage: joint model versus separate model

The main difference between these two approaches lies at the *first stage* of the IPD analysis because there are two ways in which the data can be analysed.

The first approach consists of obtaining the estimates, standard errors (SEs) (variances) that come from each univariate analysis separately for each outcome (*separate* model), and the second consists of calculating the estimates and SEs taking into account all the outcomes in a single model (joint model).

4.3.1.1 Modelling outcomes separately - separate model

Recall at the end of each study for a final score analysis, the *j*th patient in the *i*th study will provide a final score for each outcome. Each individual patient response will be referred to as $\mathbf{Z}_{ijk}$, i.e. the response value for outcome (*k*) from patient (*j*) in the *i*th study. For each outcome separately and in each trial, the general form of the UFMAs for the final score analysis is provided in Equation 4.1.

$$
\begin{aligned}
&\mathrm{z}_{ijk} = \theta_{ik} + \beta_{ik}\mathrm{T}_{ij} + \varepsilon_{ik} \\
&\textit{where} \\
&\varepsilon_{ik} \sim N(0, \sigma_{ik}^2) \qquad (4.1) \\
&\textit{and} \\
&\mathrm{cov}(\varepsilon_{i(m-1)}, \varepsilon_{im}) = 0
\end{aligned}
$$

In Equation 4.1, θ_{ik} is the control group effect (intercept) for outcome *k*, β_{ik} is the treatment effect for outcome (*k*) and study (*i*), while T_{ij} identifies the treatment group indicator for the *i*th study and the *j*th patient.

In the separate specification (Equation 4.1), $\text{cov}(\varepsilon_{i(m-1)}, \varepsilon_{im})$ is equal to *0* and thus the outcomes are considered separately in a separate model. For the motivating example (where three continuous outcomes will again be considered), the model in Equation 4.1 can therefore be fitted as three separate univariate models, where each patient (*j)* contributes to three separate follow-up responses (one for each outcome *k*), to three models containing each outcome, one at a time [76], for example:

$$\begin{aligned} z_{ij1} &= \theta_{i1} + \beta_{i1} T_{ij} + \varepsilon_{i1} \\ z_{ij2} &= \theta_{i2} + \beta_{i2} T_{ij} + \varepsilon_{i2} \\ z_{ij3} &= \theta_{i3} + \beta_{i3} T_{ij} + \varepsilon_{i3} \end{aligned} \qquad (4.2)$$

4.3.1.2 Modelling outcomes jointly - joint model

For each outcome and each trial separately, the general form of the model for the final score analysis becomes:

$$\begin{aligned} & z_{ijk} = \alpha_{ik} + \beta_{ik} T_{ij} + \varepsilon_{ik} \\ & \textit{where} \\ & \varepsilon_{ik} \sim N(0, \sigma_{ik}^2) \\ & \textit{and} \\ & \text{cov}(\varepsilon_{i(m-1)}, \varepsilon_{im}) = \sigma_{(m-1),m} \end{aligned} \qquad (4.3)$$

In the joint specification (Equation 4.3), $\text{cov}(\varepsilon_{i(m-1)}, \varepsilon_{im})$ is equal to $\sigma_{(m-1),m}$ and thus the outcomes are considered jointly in this model. For the motivating example (where three

continuous outcomes will again be considered), the model in Equation 4.3 can therefore be fitted as a joint univariate model, where each patient (*j)* contributes to three follow-up responses (one for each outcome *k*), to a single model containing all the three outcomes jointly [17], for example:

$$\begin{pmatrix} z_{ij1} \\ z_{ij2} \\ z_{ij3} \end{pmatrix} = \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \alpha_{i3} \end{pmatrix} + \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \beta_{13} \end{pmatrix} \times T_{ij} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix}$$

$$where \qquad (4.4)$$

$$\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma_i \right) \qquad \Sigma_i = \begin{pmatrix} \sigma_{i1}^2 & \sigma_{i12} & \sigma_{i13} \\ . & \sigma_{i2}^2 & \sigma_{i23} \\ . & . & \sigma_{i3}^2 \end{pmatrix}$$

It is fundamental to recall that, having calculated the covariances, for example, for outcome one and outcome two, $\sigma_{i12} = \rho_{pi12}\sigma_{i1}\sigma_{i2}$ between the patient outcome responses, and the standard errors σ_{i1} and σ_{i2}, it is possible to also calculate the *patient-level* correlations ρ_{pi12} between individual outcome responses [86] as follows:

$$\rho_{pi12} = \frac{\sigma_{i12}}{\sigma_{i1}\sigma_{i2}} \qquad (4.5).$$

4.3.2 Second stage

The first stage therefore provides treatment effect estimates and their variances for each outcome in each trial and their within-study correlations. Having estimated these parameters, it is possible to implement the UFMA model [17], obtaining the pooled-effect estimates and their standard errors and the 95% confidence intervals. It is also possible to apply and fit the MFMA model [17] and once more obtain the pooled-effect estimates, their standard errors and the 95% confidence intervals.

In section 4.3.2.1, UFMA will be described, explaining the difference between the approach that takes into account the estimates and the variances arising from the application of the *separate* model (Equation 4.1 and Equation 4.2) and the approach that considers the estimates and the variances deriving from the implementation of the joint model (Equation 4.3 and Equation 4.4).

In section 4.3.2.2, MFMA will be described, highlighting the theoretical difference with the approach that considers the estimates and the variances deriving from the implementation of the joint model.

4.3.2.1 Defining the univariate fixed-effects meta-analysis

UFMA was applied and described in Chapter 3, section 3.2.1, in Equation 3.1 and Equation 3.2. In this section the UFMA model will be recalled taking into consideration the differences between the separate model (Equation 4.1 and Equation 4.2) and the joint model (Equation 4.3 and Equation 4.4) described in the previous sections of this chapter.

For a single outcome (k), the UMFA model assumes that the obtained estimates of the treatment effect for the ith study are normally distributed about a common, fixed, true effect and variance, which are assumed to be known.

Before going on to describe the simulation study, it is necessary to introduce the difference between applying the UFMA using the joint approach, as defined in Equation 3.1 and Equation 3.2, and applying the UFMA using the separate approach. The separate approach will be defined as UFMAs. In this case, the UFMA takes into account the treatment effect estimates and their variances that come from the implementation of the joint model (Equation 4.1 and Equation 4.2) described in section 4.2.1.1. The joint approach is the model that was already described in Chapter 3 and which will be defined as UFMAj. In This case UFMA considers the treatment effect estimates and their variances that come from the

implementation of the joint model (Equation 4.3 and Equation 4.4) described in section 4.2.1.2.

4.3.2.2 Defining the difference between univariate fixed-effects meta-analysis joint approach and multivariate fixed-effects meta-analysis

Notably, the difference between UFMAj and MFMA lies in the fact that, while MFMA takes into account the estimates, SEs (variances) and the covariances that derive from the model estimated at the first stage, the UFMAj model takes into account only the estimates and the SEs and assumes that the covariances are fixed and equal to 0 as the hypothesis is that the within-study correlation between outcomes is equal to 0 [87].

As was explained in the previous chapter, at the second stage of the IPD-MA the estimates of outcome effects and variances were used to obtain the overall results for the UFMA separate and joint models and the MFMA approach [79].

4.3.2.3 Defining the multivariate fixed-effects meta-analysis model

MVMA was applied and described in Chapter 3, section 3.2.2, in Equation 3.3 and Equation 3.4. In MVMA, all outcomes are analysed simultaneously.

4.4 Missing individual participant data mechanisms

As mentioned in the previous section, there are three missing data mechanism types. In this section, the aim is to recall the missing data mechanism from the statistics literature and to relate the theory of the missing data mechanism to the scenarios considered in this simulation analysis.

i. The definition of data MCAR is that the likelihood of missing data is unrelated to any observed or unobserved outcomes. In the context of RCTs, under MCAR, the chance of missing data is the same for individuals in different treatment groups. As an example, in rheumatoid arthritis, the core outcome APR is often measured using a laboratory test. If

the test result was invalidated as a result of a laboratory error in the sample, then the data could be MCAR, as the error is equally likely to occur in any patient in the study (i.e. regardless of treatment received) [82].

ii. When the likelihood of missing data is related to observed outcomes but not to unobserved outcomes, the missing data mechanism is referred to as MAR. For example, if in a clinical trial the dropout is more likely for high values compared to low values of the outcome tested, for example, that patients with more TJC, SJC and high values of pain (evaluated on a scale 0-100), but all patients with high values of these tested outcomes have the same chance of dropout and all patients with low values of these outcomes have the same chance of dropout, the missing data mechanism is MAR [82].

iii. However, as mentioned above, if the missing data are related to the treatment, then the missing data mechanism is MNAR, because the missing data depend on the unobserved outcome data. As an example, in rheumatoid arthritis, where the disease activity outcomes are important, it is quite likely that patients assigned to the control group, taking the placebo, with the most active disease will drop out due to lack of efficacy. In the group receiving the experimental treatment, assuming there is some beneficial effect, fewer patients will drop out. This can often lead to groups with unequal sizes because fewer of the patients with severe disease remain in the control group, biasing the end result [82].

4.4.1 Description of the simulation study scenarios

This simulation study examines eight different scenarios in total, categorised (i) to (viii) (Table 4.1). Therefore, in the following table (Table 4.1) there will be a full description of these eight different scenarios, considered according to the different levels of missing data introduced and with the different levels of the assumed *patient-level* correlation ρ_p taken into account. It should be noted that missing data have been introduced for some scenarios (scenarios i and ii of MCAR and scenarios v and vi of MAR) only in one outcome (outcome 1), while for other scenarios (scenarios iii and iv of MCAR and scenarios vii and viii of MAR)

missingness was introduced in all outcomes. It also has to be considered that the amount of missing data simulated varies in the scenarios considered.

Table 4.1 Missing IPD scenarios used in the simulation

Scenario	Assumed *patient-level* correlation: ρ_p	Description
		Missing completely at random (MCAR)
i	0 0.2 0.5 0.8 RA_corr	20% MCAR in the IPD for outcome 1
ii	0 0.2 0.5 0.8 RA_corr	40% MCAR in the IPD for outcome 1
iii	0 0.2 0.5 0.8 RA_corr	20% MCAR in the IPD for all outcomes
iv	0 0.2 0.5 0.8 RA_corr	40% MCAR in the IPD for all outcomes
		Missing at Random (MAR)
v	0 0.2 0.5 0.8 RA_corr	20% MAR in the IPD for outcome 1
vi	0 0.2 0.5 0.8 RA_corr	40% MAR in the IPD for outcome 1
vii	0 0.2 0.5 0.8 RA_corr	20% MAR in the IPD for all outcomes
viii	0 0.2 0.5 0.8 RA_corr	40% MAR in the IPD for all outcomes

RA_corr: Rheumatoid Arthritis correlations from a clinical trial previously conducted, 0.67 (Y_1 & Y_2), 0.55 (Y_1 & Y_3), 0.38 (Y_2 & Y_3)

In this section, all the scenarios have been defined. In the next section, the aim is to describe how the missing data at the IPD level for each study within each simulation have been generated.

4.5 Generating the missing outcome data at the individual participant data level

As described in Chapter 3, the IPD complete case datasets were generated. In this section, the aim is to describe how the missing data in the IPD were introduced for each study in each simulation at the IPD level.

As defined in Chapter 3, for each of the three outcomes, the treatment effect is defined as the mean difference between the group of patients randomised to control (C) and the group of treated patients (T).

Firstly, the number of studies (n) within each meta-analysis and each simulation scenario was fixed at five studies and 10 studies respectively. Secondly, the goal was to simulate IPD for each study, adhering as far as possible to what had been assessed in some SRs for treatment of rheumatoid arthritis analysed. Thirdly, the ratio between the sample sizes of patients in each treatment arm was therefore set as 1:1.

The true effect estimates (mean differences) were set equal to -5 for outcome one (TJC), -3 for outcome two (SJC) and -6 for outcome three (pain). While the choice of treatment effect estimates may be seen as an arbitrary one in a simulation study, typically these were set such that they were the ‘average’ treatment effect values observed in the rheumatoid arthritis reviews. As an example, for TJC, an interpretation of a mean difference of -5 would reflect that ‘on average’ five fewer TJCs were observed in the treatment group than in the control group.

In the five study scenarios, the sample size (N) for each study ranged from 30 to 70 in each treatment arm, and this was from 10 to 100 in each treatment arm for meta-analyses containing 10 studies. Sample sizes were fixed for each study across all simulations. For scenarios where n = 5 studies were simulated, the first study contained 30 patients in each treatment arm, increasing in increments of 10 additional patients up to study 5, which had 70 patients.

The characteristics of the missing data mechanism are as follow.

i. Missing data has been introduced in one outcome or in three outcomes.

ii. The amount of 20% or 40% of missing data has been considered missing for each study.

    a. First, the scenario where *n* = 5 studies were simulated could be described as follows. In the scenario where each study had 20% missing data had been considered, the number of data deleted in each treatment group was equal to 12 for the first study, 16 for the second study, 20 for the third study, 24 for the fourth study, and finally 28 for the fifth study. In the scenario where 40% of missing data has been considered, the number of data deleted in each treatment group was equal to 24 for the first study, 32 patients for the second study, 40 for the third study, 48 for the fifth study and finally 56 for the fifth study.

    b. Second, the scenario where *n* = 10 studies were simulated will be described as follows. In each study that had 20% of missing has been considered, the number of data deleted in each treatment group was equal to 2 for the first study, 4 for the second study, 6 for the third study, 8 for the fourth study, 10 for the fifth study, 12 for the sixth study, 14 for the seventh study, 16 for the eighth study, 18 for the ninth study and finally 20 for the tenth study. In the scenario where 40% of missing data has been considered, the number of data deleted in

each treatment group was equal to 4 for the first study, 8 for the second study, 12 for the third study, 16 for the fourth study, 20 for the fifth study, 24 for the sixth study, 28 for the seventh study, 32 for the eighth study, 36 for the ninth study and finally 40 for the tenth study.

For MCAR, the probability of an observation (the outcome value for TJC, SJC or pain) being missing does not depend on observed or unobserved measurements. Therefore, the participant data were excluded from each study by randomly excluding outcome data from patients assigned to either the treatment or control group. For scenarios with 20% missing data, 20% of the data values for each study were randomly removed from treatment or control group. For scenarios with 40% missing data, 40% of the data values for each study were randomly removed from treatment or control group. On average, the number of participants with missing data was expected to be similar for both groups. The missing data were considered to be MCAR as the random selection ensured that the missingness was not dependent on either the treatment allocation or the outcome value.

For MAR, given the observed data (the outcome value for TJC, SJC or pain), the missingness mechanism does not depend on the unobserved data. Therefore, the participant data were excluded from patients assigned to either the treatment or control group, but the exclusion of data was not completely random. For scenarios with 20% missing data, all data values above the 80^{th} percentile for each study were removed, while for 40% missing data, all values above the 60^{th} percentile for each study were removed. This mechanism resembled the situation where patients might drop out because treatment is ineffective (i.e. higher values are associated with more TJC, SJC and more pain).

The detailed SAS code used to generate IPD and to introduce missing data at individual participant level in the simulated data sets is contained in Appendix D. It should be noted that a MNAR missing data mechanism was not considered in this simulation in order to maintain equal balance in the treatment groups; this is commented on in the discussion.

4.6 Assessment of performance

As defined in Chapter 3 in Section 3.9, some parameters to assess the performance of the estimates of univariate and MFMA has been calculated. Therefore as in Chapter 3 the Bias (Equation 3.9), standard error (SEsim), mean square error (MSEsim), coverage, power and BoS (Equation 3.10) were calculated.

Once the data has been simulated and analysed, the interest will be in comparing the results obtained for the complete case scenario with the results obtained for the missing data scenarios. Furthermore, the interest is focused on the comparison between the results obtained from MFMA and the UFMAj. The comparison between the results obtained from MFMA and the UFMAs is also of interest, as, finally, is the comparison between the results obtained from the UFMAj and UFMAs models.

4.7 Simulation results

Here the focus is on the results obtained from the analysis of IPD under assumption of MCAR and MAR.

When zero *patient-level* correlation ρ_p was considered, unrelated to the missing data mechanism, performance indicators for MFMA and UFMA (joint and separate UFMAj and UFMAs approaches) were identical for all scenarios considered (Table E.2 to Table E.9 in Appendix E). In the tables of the appendix E, μ_1 refers to the outcome 1 (TJC), μ_2 refers to the outcome 2 (SJC) and finally μ_3 refers to the outcome 3 (pain).

4.7.1 The complete case scenario

For the complete case (results from Chapter 3 for comparison only and considering only UFMAj; Table E.1 of Appendix E), there was a small bias for all the three outcomes simulated and for all scenarios considered, with only a small notable benefit to be gained from using MFMA over UFMAj for all three outcomes when the *patient-level* correlation was high (ρ_p = 0.8). For example, when n=5 studies were simulated the bias for outcome 1 was equal to 0.035, for outcome 2 it was equal to 0.020 and for outcome 3 it was equal to 0.023 when the MFMA model was applied, while it was equal to 0.044 for outcome 1, to 0.022 for outcome 2 and 0.033 for outcome 3 when UFMAj was implemented. As expected, there were gains in precision (smaller SEs) for all scenarios when MFMA was applied over UFMAj; this also translated into improved MSEs for all scenarios. As the *patient-level* correlations increased, there was also an increase in the %BoS for all outcomes. As an example, for five studies the %BoS for the three outcomes respectively was 1.8%, 1.9% and 1.7% when the *patient-level* correlation (ρ_p) was 0.2, compared to 2.8%, 2.9% and 2.9% for the three outcomes when the *patient-level* correlation (ρ_p) was 0.8. The coverages estimated were around the nominal range for all scenarios and methods of estimation.

4.7.2 Missing completely at random

When MCAR missingness was introduced in one outcome (outcome 1) and the percentage of missingness was at its lowest level (20%) (Table E.2), there were elevated biases in the pooled effect for outcome 1 with missing data, particularly when five studies were considered. The introduction of MCAR also caused slightly raised biases in outcomes with no missing data, particularly with higher *patient-level* correlations with five study scenarios. This bias appeared to be less marked when the number of studies increased to 10. For the outcome with missing data, there appeared to be no benefit (in terms of bias reduction) in the MFMA approach and the least-biased estimate appeared to be UFMAs. For example, for the scenario where the number of simulated studies was five and the *patient-level*

correlation (ρ_p) was equal to 0.5, the bias was equal to 0.082 when the MFMA model was applied, while it was equal to 0.013 when the UFMAs was applied. Increasing the number of studies to 10 and keeping the same value of *patient-level* correlation (ρ_p), the bias calculated was equal to 0.041 when the MFMA model was applied while it was equal to 0.035 when UFMAs was implemented. When increasing the *patient-level* correlation (ρ_p) to 0.8 when n=5 studies were simulated, the bias was equal to 0.084 when the MFMA model was applied, while it was equal to 0.012 when the UFMAs was applied. Increasing the number of studies to 10 and keeping the same value of *patient-level* correlation (ρ_p = 0.8), the bias calculated was equal to 0.039 when the MFMA model was applied, while it was equal to 0.035 when UFMAs was implemented. In the two outcomes with no missing data, there were trade-offs to be made: UFMAs often had larger biases in the estimation of outcome three and sometimes MFMA was preferred but there was no obvious pattern in the consistency of results.

Nevertheless, there was a visible advantage in MFMA when observing the BoS. In outcome one, where there were missing data, greater percentage changes in BoS were observed, providing evidence of BoS when using the MFMA approach. Larger %BoS changes were observed between MFMA and UFMAs than MFMA and UFMAj, and, as previously mentioned, this was more pronounced when there were missing data (in this case outcome one only) and the *patient-level* correlation (ρ_p) increased. As an example, for 10 studies with a *patient-level* correlation (ρ_p) of 0.8, the %BoS for the three outcomes respectively between MFMA and UFMAs was 17.0%, 4.4% and 4.1% compared to 4.1%, 4.3% and 3.9% between MFMA and UFMAj. In another example, the %BoS for MFMA vs. UFMAs using a *patient-level* correlation (ρ_p) of 0.2 (10 studies) was 4.6%, 2.7% and 2.5% for the three outcomes compared to the 17.0%, 4.4% and 4.1% for a *patient-level* correlation (ρ_p) of 0.8.

As expected, the precision was always better for the MFMA approach over both UFMAs and UFMAj. The precision increased as the *patient-level* correlation (ρ_p) increased. Notably, the precision for UFMAj was always at least the same (but mostly improved) over UFMAs. The same gains were not observed in the outcomes with no missing data. Overall, this provided an interesting result in terms of the MSEs. For all scenarios and outcomes, MFMA was shown to be improved over UFMAs and UFMAj and UFMAj was shown to be improved over UFMAs.

Coverage was actually slightly lower than the nominal for MFMA than both the UFMA approaches but the difference was small and not thought to be of concern.

4.7.2.1 Missing completely at random - doubling the amount of missingness

When the level of missingness doubled to 40% but was still only introduced in one outcome (Table E.3), the results were not too dissimilar from a reduced level of missingness discussed in section 4.9.2. In fact, the amount of bias did not increase (and in some cases even decreased). This is a not an atypical finding when the missingness is MCAR, as the missingness is unrelated to the observed or unobserved outcome data. There was a marginal gain in using MFMA over the two univariate approaches when using 10 studies and a *patient-level* correlation (ρ_p) set to 0.8. In the outcome with the missing data (outcome 1), there were raised SEs as a result of the inefficiency of having less data available for analysis but the gains remained for MFMA. For all scenarios, on average MFMA was still shown to be superior when considering MSEs, The increase in missing data also encouraged extra BoS from the MFMA approach. As an example, considering 10 studies with a *patient-level* correlation (ρ_p) of 0.8 (comparing to UFMAs), the %BoS for 20% missing data in outcome one was 17% compared to 30.6% with 40% missingness. There were no new notable concerns with coverage.

4.7.2.2 Missing completely at random – missingness in all outcomes

Tables E.4 and E.5 show the results for 20% missingness and 40% missingness respectively when missingness was introduced in all three outcomes. The results were mainly similar for missingness in one outcome but the general trend could be applied to all three outcomes. For example, there was now an increase in SEs for all three outcomes as a result of introducing missingness in each outcome. The %BoS was also consistent for all outcomes. As an example, for 40% missingness in all outcomes, for 10 studies with a *patient-level* correlation (ρ_p) set to 0.8, the %BoS for each outcome was 33.4%, 34.9% and 33.0% respectively. The same level of borrowed strength was only observed in outcome one when missingness was introduced only in this outcome. For both levels of missingness, in terms of MSEs, MFMA was still preferred over the two univariate approaches for all scenarios and the percentage gains were larger when there were more missing data and the *patient-level* correlations were higher. One notable concern was the MFMA coverage when there was 40% missing data in all outcomes when 10 studies were considered. This is likely caused by the trade-off between a slightly increased bias (as a result of the missingness) in the estimation of the parameter estimates and the increased precision in the estimation of standard errors. This issue was not evident when there were lower levels of missingness (20%).

4.7.3 Missing at random

When a MAR missing data mechanism was introduced in one outcome (outcome 1) and the percentage of missingness was at its lowest level (20%) (Table E.6), there were much larger biases observed in the outcome with missing data than the MCAR scenarios. In the majority of cases when there were missing MAR data, MFMA outperformed both UFMAs and UFMAj in terms of bias reduction. There was just one example when this was not the case, but this was negligible (a difference in bias of 0.01).

There was evidence that the bias in UFMAj and MFMA reduces as the *patient-level* correlation (ρ_p) increases, but no such reductions were observed for UFMAs. As an example, for 10 studies, when the correlation was set to 0.2, the average bias in the MFMA estimate was 0.341 compared to 0.277 for a correlation of 0.8, nearly a 20% reduction. There were no obvious benefits of using MFMA in terms of bias reduction in the two outcomes where there were no missing data.

There were benefits of increased precision (lower average SEs) for all scenarios using MFMA over the two univariate approaches and, as before, this translated into improved MSEs. Again, larger BoS was observed when comparing MFMA with UFMAs than with UFMAj, although the amount of BoS was not as pronounced as for MCAR. The improvement in increasing the *patient-level* correlation (ρ_p) was not observed as before, with the BoS remaining consistently around 4-5% when comparing MFMA with UFMAs (10 studies). There was a slight drop in coverage for the outcome with missing data, which is to be expected, but this largely affected all three methods of estimation in the same way.

4.7.3.1 Missing at random – doubling the amount of missingness

As the amount of missingness increased to 40%, larger biases were observed in the outcome with missing data and again MFMA appeared to outperform UFMAs and UFMAj in all but one case in terms of bias reduction (Table E.7). Again, with the one case, the bias change was only 0.02 (with a marginal benefit of UFMAj only) and was not considered problematic. As the *patient-level* correlation (ρ_p) increased, the difference in the bias between MFMA and UFMAs in particular was now quite dramatic. As an example, for five studies with a *patient-level* correlation (ρ_p) set to 0.8, the MFMA bias estimate was 0.288 compared to 0.592 for UFMAs, a 51% reduction in bias. As previously mentioned, the UFMAs approach does not improve as the *patient-level* correlation (ρ_p) increases. Benefits in reduced SEs and MSEs in favour of MFMA were seen throughout. A similar pattern in the consistency of lower BoS was again observed when compared to MCAR with

the same levels of missingness. Increasing the *patient-level* correlations also did not appear to consistently increase the level of BoS as with MCAR, but larger BoS was observed than MAR with only 20% missingness. Coverages were affected more by having an increased level of MAR data in outcome 1, with no benefit for any of the estimation methods when considering this performance parameter, although there is an argument that UFMAj did perhaps perform slightly better.

4.7.3.2 Missing at random – missingness in all outcomes

Tables E.8 and E.9 show the results for 20% missingness and 40% missingness respectively when missingness was introduced in all three outcomes. More often than not, MFMA outperformed UFMAs and UFMAj in terms of bias reduction and again SEs and MSEs always benefitted MFMA. Again, there was consistency in the amount of BoS across all three outcomes for the same scenario with higher levels of BoS as the amount of missingness increased from 20% to 40%. Coverage was lowered for all outcomes for all scenarios based on the level of missingness and the MAR mechanism. There was some suggestion that UFMAj was the better of the estimation methods when considering coverage, although the benefit was not considerable.

4.8 Discussion

Various forms of missing data plague the validity of conclusions to many meta-analyses. Adjustment approaches are available to systematic reviewers to adjust for missing data at the study level, whether this be for outcome reporting bias or publication bias. As an example, the MVMA approach proposed in Chapter 3 (alongside other methods, [88], for example the Bias bound estimate) has been proposed as an adjustment method for outcome reporting bias and, for example, Egger, trim and fill [89] are available methods for adjusting for publication bias. However, little is known about the impact of missing data within the IPD in a trial on the conclusions when reviewing meta-analyses, or how reviewers might adjust for this form of missingness, particularly if the amount of missing participant

data is unknown, because it is not reported in the trial publication and the IPD are not available. In some cases, trial authors may adjust for this form of missing data in their primary analyses, and therefore reviewers may be able to look at the sensitivity on the pooled results using reported, unadjusted and adjusted analyses. However, the literature suggests that trialists often do not impute missing data [85] and, even when they do, in a large majority of cases, an inappropriate method was used. Most reviewers will only assess this form of missingness in the risk of bias assessment, but often the lack of quality of available evidence is not accounted for in the overall review conclusions.

Assuming the IPD are available, reviewers could apply an appropriate imputation method to the data before synthesising the data, and similarly could look at the sensitivity of the pooled results from the imputed data to the complete case data. However, IPD are rarely available for all studies and applying some of the advanced imputation methods may be beyond the skill set of some reviewers. Moreover, determining the missing data mechanism, and hence selecting an appropriate imputation strategy, can be difficult to determine, especially without any first-hand knowledge of the way the trial was conducted.

In a similar way that MVMA offers advantages in the presence of missing study data, the concept that the method may also improve pooled estimates in the presence of missing participant data has been explored in this chapter. In a similar set-up to the previous chapter, a two-stage IPD-MA simulation study was employed such that missing participant data could be incorporated into the trial data under both a MCAR and a MAR missing data mechanism. The two-stage approach was particularly advantageous here, as insight into the benefit of the MVMA approach over the two univariate specifications in the second stage under this type of missingness could also be observed, if reviewers were only presented with AD from the trial reports (recalling the two-stage approach essentially reduces the IPD to AD for the second stage).

The results from the simulation study were encouraging and, considering the mean square error (MSE) as a trade-off between preferred bias and improved precision, then the MVMA

approach was preferred over the two different univariate specifications irrespective of an MCAR or MAR missing data mechanism and two different levels of missing data (20% and 40%). The benefit in MFMA when looking at %BoS was also clear to see, particularly as the amount of missing data and the *patient-level* correlation (ρ_p) between outcomes increase. More BoS occurred in outcomes when the missingness was MCAR than MAR. Whilst some very minor advantages were seen in some performance indicators from the univariate approaches (coverage and bias in some cases), the differences in most cases were small and simulation error could not be ruled out. As a consequence, the general recommendation would be that the multivariate approach would still be preferred.

A limitation of this simulation study is that a MNAR missing data mechanism within the IPD was not generated. This was omitted by design to prevent an imbalance in the number of participants within each treatment arm when the missingness had been introduced. Nevertheless, outcome reporting bias that was investigated in Chapter 3 is an MNAR missing data mechanism, and therefore it is anticipated (as shown in Chapter 3) that a MVMA would still show the same benefits if MNAR was introduced in the IPD.

Overall, the MVMA approach offers an improved estimate of pooled-effects estimate for a range of different missing data mechanisms of varying degrees of missingness in the IPD. This is particularly beneficial when the true missing data mechanism and the amount of missingness is unknown. The method also offers the advantage that it could be simpler for reviewers to implement over imputation methods, if the IPD are available.

Chapter 5 – Assessing the impact of multivariate meta-analysis using Systematic Reviews of Rheumatoid Arthritis

5.1 Introduction

In Chapter 2, an assessment of ORB was carried out through a set of SRs for the pharmacological treatment of RA. The process identified high and low risk of bias for trials where selective non-reporting was suspected.

In Chapter 3, a simulation study was presented where both UFMA and MFMA models were applied to IPD for a number of different scenarios. The results from Chapter 3 demonstrated that there were benefits in terms of improvements in the performance indicators when applying the MVMA approach, particularly in the presence of suspected ORB.

In this current chapter, the aim is to consider a selection of these SRs presented in detail in Chapter 2 and to determine the statistical impact (and where possible the potential clinical impact) on the conclusions to the pooled effect estimates when a MVMA approach is applied over standard UVMA, when ORB is suspected.

In this chapter, the multivariate random-effects meta-analysis (MRMA) model is also introduced for the first time as for some of the examples, a REMA may be more appropriate. The results of each application are presented using both fixed and random effects approaches for comparison. To assist with the interpretation, a multivariate measure of heterogeneity is introduced and also a multivariate forest plot is presented.

5.2 Review eligibility criteria for assessment

The 21 reviews containing 172 trials that were considered for ORB assessment in Chapter 2 were also considered here for further analysis.

There were three eligibility criteria according to which the SRs were considered and analysed in this chapter:

- The SR had to contain at least three clinical trials. The reason for this choice is that it would be difficult to fit the MVMA model if there were fewer studies. This is due to the fact that in the SRs that have been assessed the MA is conducted with AD.

- The SR had to have considered the individual core outcomes as review outcomes of interest, rather than a composite measure of all the core outcomes.

- The SR had to consider only one MA for each of the core outcomes taken in account.

On the basis of these eligibility criteria, among the 21 reviews considered only seven could be taken further and analysed. Among the 14 SRs that were not analysed:

- *Four* SRs (*Azathioprine* [47], *Cyclophosphamide* [49], *Cyclosporine* [50], and *Injectable gold* [51]) were not considered because they analysed a limited number of clinical studies.

- Ten SRs (*Methotrexate monotherapy versus methotrexate combination therapy with non-biologic disease modifying antirheumatic drugs for rheumatoid arthritis* [41], *Penicillamine* [53], *Folic acid and folinic acid for reducing side effects in patients receiving methotrexate* [56], *Anakinra* [59], *Certolizumab Pegol* [60], *Etanercept* [61], *Infliximab* [62], *Golimumab* [63], *Tocilizumab* [64], *Glucocorticoids* [65]) were

excluded because even if they were presenting more than one study for each of the core outcome, each study was measuring and analysing the outcomes at different time-points.

Seven SRs (*Antimalarials* [46], *Auranofin* [48], Methotrexate [52], *Sulfasalazine* [54], *Leflunomide* [55], *Adalimumab* [57] and *Abatacept* [58]) met the eligibility criteria and are considered in this chapter. These were the straightforward reviews where UVMA and MVMA could easily be applied. Given these reasons, it is essential to recall that in the following examples only the results in terms of MD have been considered, extracted from the SRs that were comparable with each other. The included SRs have been considered to have homogeneous measures in terms of treatment/placebo dose administered to two groups of patients.

The outcome RD was not considered in the analysis for two reasons, firstly this outcome was not meta-analysed in any of the SRs considered in this analysis, secondly RD is analysed only in RCT that have a duration greater than or equal to 48 weeks, meaning that there was no OMERACT recommendation to measure this outcome in most RCTs (89%) included in the SRs reported in this chapter. Therefore, only the remaining seven outcomes have been considered for each SR analysed. In the following section, the ORB tables will be presented for each SR analysed with the UVMA and MVMA and presented in this chapter [35].

In some SRs included in the analysis for this chapter, some outcomes were classified as FR. This happened because these outcomes were fully assessed in the original studies included in each SR but the SR itself was not considering the measurement instrument used. Therefore, although this missing data could not be classified as a suspected case of ORB, it could not be included in the MA. All these cases have been considered as missing data when the meta-analysis models were fitted even if they were not suspected of ORB.

5.3 Models applied to the systematic reviews analysed

To illustrate the impact that ORB may have in an SR containing multiple outcomes, a MVMA approach was applied to a review comparing each treatment of interest (T) with a control (C) for rheumatoid arthritis. The aim was to examine whether the summary results and conclusions from the MVMA differed to those from the original MAs performed by the review authors. The impact was assessed in terms of the change in the treatment effect estimates, change in the statistical significance of the treatment effect estimates, and change in the precision of the treatment effect estimates, for each of the outcomes of interest. All examples that will be considered in this chapter have been analysed using FEMA and REMA. The rationale for this is that in a univariate analysis, authors may select a different method for each outcome depending on levels of heterogeneity [17].

Furthermore, it is important to say that the analysis performed in this chapter is an AD-MA, as this was the approach the original authors used in the practical examples and it was impossible to obtain the IPD from the researchers for all studies. In the IPD simulations that were presented in Chapter 3 and Chapter 4, a 2-stage IPD-MA approach was used. It could be demonstrated that the AD-MA models applied in this chapter are the same models that were fitted at the second stage of the IPD-MA. In this chapter, for all examples analysed, the estimate of the MD and the estimate of the SE for each study have been used to run the models. These data are extracted from the original SRs conducted by the authors.

Firstly, UFMA and MFMA models were fitted. Secondly, URMA and MRMA models were then fitted for comparison, on the basis that the review commented that in some cases heterogeneity existed, as assessed using (I^2).

5.3.1 Univariate fixed-effects meta-analysis

In FEMA it is assumed that that all studies are estimating the same true effect. It is also assumed that the variability between study results is due solely to the sample of patients

within each study. Another important assumption is that precision depends mainly on study size. The UFMA was presented in Chapter 3 in Equation 3.1 and in Equation 3.2.

As defined in Chapter 3, the assumption for the UFMA model is that there are *m* outcomes for each study *i*. Each study *i* supplies a set of *m* estimates of the MD, $(y_{i1}, y_{i2}, ...y_{im})$ and associated variances $(s_{i1}^2, s_{i2}^2, ...s_{im}^2)$, for each outcome [16].

However, in Chapter 3, the UFMA model was applied at the second stage of the IPD-MA, while in this part of the study the UFMA is applied directly to the AD provided for each study of the SRs analysed.

5.3.2 Univariate random-effects meta-analysis

In REMA the treatment effects for the individual studies are assumed to vary around some overall average treatment effect μ.

Suppose that there are *m* outcomes for each study *i*. Each study *i* supplies a set of *m* estimates $(y_{i1}, y_{i2}, ...y_{im})$ for all the outcomes. Each summary statistic is assumed to be an estimate of a true value μ_i in each study i, and in a hierarchical structure each μ_i is assumed to be drawn from a normal distribution with mean value β and between study variance τ_i^2. Then the URMA can be specified as [73]:

$$\begin{cases} y_{i1} \sim N(\mu_{i1}, s_{i1}^2) & \mu_{i1} \sim N(\beta_1, \tau_1^2) \\ y_{i2} \sim N(\mu_{i2}, s_{i2}^2) & \mu_{i2} \sim N(\beta_2, \tau_2^2) \\ y_{im} \sim N(\mu_{im}, s_{im}^2) & \mu_{im} \sim N(\beta_m, \tau_m^2) \end{cases} \qquad (5.1)$$

where $(s_{i1}^2, s_{i2}^2, ...s_{im}^2)$ are the w/s variances and $(\tau_{i1}^2, \tau_{i2}^2, ...\tau_{im}^2)$ between-study variances. All outcomes are considered independent in the equation. To estimate the parameters of each model we used the REML method.

5.3.3 Multivariate fixed-effects meta-analysis

For MFMA, all outcomes are analysed simultaneously [73]. In addition to the two quantities required for a UFMA from each study (treatment effect estimate and the w/s variance), the w/s correlation is also required. The MFMA model was presented in Chapter 3 in Equation 3.3 and in Equation 3.4.

As defined in Chapter 3, the vector for the *m* treatment effect estimates $\mathbf{Y}_i$ $(y_{i1}, y_{i2}, ..., y_{im})$, the vector $\boldsymbol{\mu}$ $(\mu_1, \mu_2, ..., \mu_m)$ of the fixed true treatment effects for the outcomes and the w/s variance-covariance matrix $(\mathbf{S_i})$ of the effects' estimates need to be known.

However, in Chapter 3, the MFMA model was applied at the second stage of the IPD-MA, while in this part of the study the MFMA is applied directly to the AD provided for each study of the SR analysed.

5.3.4 Multivariate random-effects meta-analysis

In addition, for MRMA, all outcomes are analysed simultaneously [78]. The MRMA model is based on treatment effect estimates and has a hierarchical structure with multivariate normal distributions at each of two levels, corresponding to w/s and between-study components. Therefore, the multivariate random-effects model is presented in Equation 5.2.

$$\begin{pmatrix} y_{i1} \\ y_{i1} \\ \vdots \\ y_{im} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{im} \end{pmatrix}, \mathbf{S}_i\right) \qquad \mathbf{S}_i = \begin{pmatrix} s_{i1}^2 & s_{i12} & \cdots & s_{i1m} \\ . & s_{i2}^2 & \cdots & s_{i2m} \\ \vdots & \vdots & \ddots & \vdots \\ . & . & \cdots & s_{im}^2 \end{pmatrix}$$

$$\begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{im} \end{pmatrix}, \sim N\left(\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}, \Omega\right) \qquad \Omega = \begin{pmatrix} \tau_{i1}^2 & \tau_{i12} & \cdots & \tau_{i1m} \\ . & \tau_{i2}^2 & \cdots & \tau_{i2m} \\ \vdots & \vdots & \ddots & \vdots \\ . & . & \cdots & \tau_{im}^2 \end{pmatrix} \qquad (5.2).$$

Where $\mathcal{S}_i$ and Ω are the w/s and the between-study variance-covariance matrices respectively.

The objective of a MRMA is to estimate the mean treatment effects across studies, $(\mu_{i1}, \mu_{i2}, ..., \mu_{im})$, and the between-study covariance matrix, Ω. The model parameters were estimated using the ‘‘mvmeta’’ module [78] in Stata using the method of ML for FEMA models and REML method for REMA models.

In Section 5.3.5 the method proposed by Wei and Higgins [74] to derive and calculate the w/s covariance will be described.

5.3.5 Estimating the within-study covariance

The methods described in Section 5.3.3 and 5.3.4 assume that the w/s variance-covariance matrix is known. The entries of the w/s variance-covariance matrix could be derived straightforwardly in the simulations studies described in Chapter 3 and in Chapter 4 as IPD has been generated. In the examples described and analysed in this chapter for each SR only the AD were known. And in a situation of AD MA it is not straightforward to obtain the quantities needed for the variance-covariance matrix.

Previous studies have stated that these quantities are rarely reported or even calculated for each study. Kirkham et al. [15] describes three possible methods of obtaining the correlation between these estimates. First, the availability of IPD would allow us to calculate the w/s correlation between the estimators directly in each study (mentioned above) [73]. Second, if no IPD data are available, it may still be possible to approximate the w/s correlation using biological reasoning or expert opinion [73]. Third, Pearson correlation method can be used when it is impossible to obtain w/s correlation from IPD or from expert opinion [73].

In the examples in this chapter, the first method was followed. The *patient-level* Pearson correlations ρ_p between outcomes were provided in a previous study conducted by Professor Dr. George Wells (unpublished data). In this study, the researchers performed an individual patient data (IPD) analysis for two rheumatoid arthritis trials which recorded data on all core outcomes. These *individual-level* correlations from the two separate trials analysed were averaged and these estimates were used to approximate the likely *patient-level* correlation ρ_p estimates in other studies that did not provide them.

Therefore, the *patient-level* Pearson *correlations* needed to be taken in account and converted to obtain the w/s covariances needed to fit the MVMA model; the appropriate approximation was applied using Wei and Higgins' method [74].

Given a known *patient-level correlation* coefficient ρ_p between the two outcomes themselves, and assuming it is the same in both treatment groups, the following equation gives the analytical form for the covariance between MDs (MD_1, MD2):

$$\operatorname{cov}(MD_1, MD_2) = \sigma_{12} = \frac{n_{12t}}{n_{1t} n_{2t}} \rho_p s_{1t} s_{2t} + \frac{n_{12c}}{n_{1c} n_{2c}} \rho_p s_{1c} s_{2c} \qquad (5.3)\,.$$

In Equation 5.3, n_{1t} is defined as the number of participants reporting outcome 1 in the treatment group; n_{2t} as the number of participants reporting outcome 2 in the treatment

group; n_{12t} the number of participants reporting both outcome 1 and outcome 2 in the treatment group; and n_{12c} , n_{1c} and n_{2c} are defined in the similar way for the control group. Furthermore, in Equation 5.5, s_{1t} is defined as the standard error in outcome 1 for the treatment group, while s_{2t} is the standard error in outcome 2 for the treatment group; s_{1c} and s_{2c} are defined in the similar way for the control group. Once the w/s covariance has been obtained and the standard errors are known, it is possible to obtain the w/s correlation for outcomes 1 and 2, applying the following formula:

$$\rho_{w12} = \frac{\sigma_{12}}{s_1 s_2} \qquad (5.4)\,.$$

5.3.6 Methods to assess heterogeneity in a meta-analysis

In the previous sections, different techniques for meta-analysis have been described.

One of the main concepts of the REMA is related to the between-study variance. Therefore, if the between-study variance is assumed to be zero, then the model described in Equation 5.5 is referred to as the FEMA model. It could be said that FEMA models are sometimes preferable because they simplify the interpretation of the estimates and make the computation of the estimates easier. Nevertheless, the assumption of no between-study variation seems generally unlikely, unless it is known that the studies are performed in the same way and involve individuals sampled from the same population [90].

In order to understand if it is feasible to fit a FEMA model and to help choose the most appropriate model (fixed or random), it is useful to observe and calculate the level of heterogeneity. If the heterogeneity is calculated to be high across the studies, then it is more advisable to apply a REMA model that to apply a FEMA model. Therefore, a test for the

presence of heterogeneity needs to be defined and calculated and presented with the results of a meta-analysis.

5.3.6.1 Univariate meta-analysis scenario

A test for heterogeneity examines the null hypothesis that all studies are evaluating the same effect. The usual test statistic (Cochran's Q) is computed by summing the squared deviations of each study's estimate from the overall meta-analytic estimate, weighting each study's contribution in the same manner as in the meta-analysis [11]. The weights are defined with for the ith study. Therefore, the Cochran's equation for testing the heterogeneity is:

$$Q = \sum w_i (\mu_i - \hat{\mu})^2 \qquad (5.5)$$

Given this result, Higgins and Thompson defined [11] the index H^2 obtained with the following equation:

$$H^2 = \frac{Q}{(n-1)} \qquad (5.6)\,.$$

Using this result, Higgins and Thompson calculated the I^2 index to test the heterogeneity in the UVMA models [11].

$$I^2 = \left(\frac{H^2 - 1}{H^2} \right) \times 100\% \qquad (5.7)\,.$$

In the univariate case, this index describes the heterogeneity due to the variation across the studies considered in the meta-analysis. Therefore, the index I^2 provides an estimate of the percentage of variability in the meta-analysis due to the difference between each study.

Higgins and Thompson also suggested a simple categorisation of I^2 values, tentatively assigning adjectives of low, moderate, and high to I^2 values of 25%, 50%, and 75%. Nevertheless, one of the limits of this index in defining heterogeneity is the over-interpretation caused by the overlapping of the intervals for I^2 [11]. To avoid this misinterpretation, Higgins and Thompson defined another index, called R, and which is calculated by the following formula:

$$R = \frac{V_R}{V_F} \qquad (5.8)\,.$$

In Equation 5.8, V_R and V_F are defined as the length of the confidence intervals for the treatment effect that come from the REMA and the FEMA model, respectively.

In this section, the univariate index I^2 proposed by Higgins and Thompson has been described, but they also generalised its application to the MVMA scenarios. In the next section, the extension to the MVMA model will be described and discussed.

5.3.6.2 Multivariate meta-analysis scenario

In this section, MVMA heterogeneity will be considered. In the MVMA scenario there are multiple outcomes denoted with the letter *m*. In the examples analysed and reported in this chapter the core outcomes considered are *m*=7.

Therefore, the R calculated before for the UVMA case is applied to the MVMA scenarios as follows:

$$R = \left(\frac{V_R}{V_F} \right)^{1/m} \qquad (5.9)$$

Higgins and Thompson [11] also provided and calculated I_R^2, taking into account this new index calculated for the multivariate case. The following formula will provide the desired index:

$$I_R^2 = \left(\frac{R^2 - 1}{R^2} \right) \times 100\% \qquad (5.10).$$

This index is the multivariate counterpart of the I^2 index calculated to estimate the heterogeneity of a meta-analysis. The index I_R^2 provides the percentage of variability in the meta-analysis due to the between-study variability across the studies considered. Therefore, it is also possible to consider for index I_R^2 a simple categorisation of its values, assigning low, moderate, and high to values of 25%, 50%, and 75% respectively [90].

5.3.7 Univariate and multivariate meta-analysis: forest plot

In this section, forest plots for univariate and MVMA will be presented alongside the tables with the estimates and the results obtained from the models fitted to each single example examined. The forest plot has the following common characteristics that make it a useful tool to give a graphic view of the results of a meta-analysis:

- Each individual study is represented on a common scale.

- Each study's effect and respective confidence interval are plotted on one set of axes.

- The effect estimate (in this example the MD) is represented by a square and the size of the square is related to the weight that the study has in the meta-analysis. Therefore, it could be stated that a small and slightly informative clinical trial will

have a large confidence interval, indicating that the estimate of the treatment effect is not precise.

- The pooled MD estimate is plotted at the bottom of the graph and is usually represented as a diamond.

- The pooled estimates obtained from the UVMA model and from the MVMA model fitted with the data from each single study are represented in the forest plots. This happens both for FEMA and REMA approaches. The centre of the diamond represents the pooled estimate and the width of the diamond shows the confidence interval.

- The forest plot also allows a graphical examination of the degree of heterogeneity between studies. The more confidence intervals are overlapping, the less heterogeneity could be detected by the meta-analysis of the SR.

- The forest plots present for each outcome considered both UVMA and MVMA estimate (diamond). While the global UVMA estimate is obtained averaging all the results from each single study for each single outcome, the global MVMA estimate is obtained considering simultaneously all the outcomes considered in the MA.

To obtain the forest plots presented, the STATA command mvmeta_forest ‘version 10.1’ was used.

5.3.8 Parameters used to assess the difference between univariate meta-analysis and multivariate meta-analysis

In analysing the results, which are presented in section 5.4 of this chapter, the aim was to assess the difference between the univariate model (UFMA or URMA) and the multivariate model (MFMA or MRMA).

To assess the difference between the models fitted in this analysis the attention was focused on discussing:

- the change in the estimates of the pooled MD obtained from fitting the univariate model (UFMA or URMA) and the estimates of the pooled MD obtained fitting the multivariate model (MFMA or MRMA);

- the change in the estimates of the Standard Error (SE) obtained from fitting the univariate model (UFMA or URMA) and the estimates of the MD obtained fitting the multivariate model (MFMA or MRMA);

- the change in the P-values obtained from fitting the univariate model (UFMA or URMA) and the estimates of the MD obtained from fitting the multivariate model (MFMA or MRMA);

- the result of the heterogeneity index (I^2 UVMA) provided from the original SR;

- the heterogeneity index (I_R^2 MVMA).

In the following section, the main results for each SR and for each outcome considered will be presented, focusing attention on the comparison between UVMA and the MVMA model. The interest will be addressed also to the comparison between REMA and the FEMA approach.

5.4 Results

For each SR, a summary of the ORB assessment will be reported, a summary of the conclusions from the actual review will be presented, and finally a description of the main difference in results between univariate and multivariate will be explained.

5.4.1 Antimalarials for treating rheumatoid arthritis

5.4.1.1 Summary of the outcome reporting bias assessment

The SR 'Antimalarials for treating rheumatoid arthritis' was conducted by Suarez-Almazor et al. [46]. This SR presents the results of UFMA.

The objective of this SR was to compare the short-term efficacy and toxicity of Antimalarials for the treatment of RA by comparing hydroxychloroquine with placebo. The review aimed to synthesise data on all of the RA core outcomes, amongst a selection of other outcomes. This SR identified four eligible studies, with a combined total of 300 randomised to hydroxychloroquine and 292 to placebo. No studies were excluded due to 'no relevant outcome data'.

The results for the ORB assessment for the core outcomes (undertaken for this review in Chapter 2) are summarised in Table 5.1.

Table 5.1 ORB matrix for Antimalarials systematic review

Study	TJC	SJC	Pain	Phy. Global	Pat. Global	Function	APR
Davis '91	✓	✓	×(D)	×(H)	×(H)	×(FR)	✓
Clark '93	✓	✓	✓	×(C)	×(C)	×(FR)	✓
Blackburn '95	✓	✓	✓	✓	✓	×(D)	✓
HERA '95	✓	✓	✓	✓	✓	✓	✓

TJC: Tender Joint Count; SJC: Swollen Joint Count; Phy.: Physician; Pat.: Patient; APR: Acute Phase Reactant; FR: Fully reported
ORB Classification [9] A: Partial reporting, High Risk ORB; B: Partial reporting, No Risk ORB; C: Partial reporting, Low Risk ORB; D: None, High Risk ORB; E: None, High Risk ORB; F: None, Low Risk ORB; G: None, High Risk ORB; H: None, Low Risk ORB; I: NA, No Risk ORB.

Only the study HERA '95 reported on all of the core outcomes, while the core outcomes TJC, SJC and APR were reported in all four studies. Amongst the other outcomes, there was a mixture of high and low ORB classifications (justifications for these classifications can be found in Appendix A (from Chapter 2)). Function was classified as FR in two studies (Davis '91 and Clark '93) because they reported on a measure of function in the trial report, but not using an acceptable measurement instrument for inclusion in the review.

5.4.1.2 Summary of the results from the original systematic review

In this particular SR, the original study authors applied a FEMA to all seven core outcomes being considered. For TJC (0%), SJC (0%), Phy.Global (0%) and APR (0%) the value of the heterogeneity index I^2 was low. For other outcomes as pain (88%) and function (100%) the value of heterogeneity index was high. For the outcome Pat.Global the heterogeneity index was moderate and equal to 51%. The pooled results for each outcome, as reported in the original review, can be found in Table 5.2 (UFMA). The use of hydroxychloroquine appears to be efficacious (the pooled estimates for the MD is negative) for the treatment of RA when considering all other core outcomes. For the outcome function the pooled MD estimate is equal to -0.06. All the estimates of the MD are statistically significant, with the exception of the outcome function (p-value = 0.609).

Table 5.2 Meta-analysis results for the Antimalarials data set

Outcome	Univariate meta-analysis				Multivariate meta-analysis			
	Mean difference	SE	P-value	95% CI	Mean difference	SE	P-value	95% CI
	UFMA				**MFMA**			
TJC	-2.57†	0.62	<0.001***	-3.78; -1.36†	-1.44	0.54	0.007**	-2.50; -0.39
SJC	-3.71†	0.59	<0.001***	-4.86; -2.57†	-2.84	0.53	<0.001***	-3.87; -1.80
Pain	-0.45†	0.14	0.001**	-0.72; -0.18†	-0.12	0.06	0.049	-0.23; -0.001
Phy.Global	-0.39†	0.09	<0.001***	-0.57; -0.21†	-0.23	0.06	<0.001***	-0.34; -0.12
Pat.Global	-0.34†	0.10	0.001**	-0.53; -0.15†	-0.08	0.05	0.080	-0.17; 0.01
Function	-0.06†	0.12	0.609	-0.29; 0.17†	0.10	0.09	0.256	-0.07; 0.28
APR	-6.38†	1.09	<0.001***	-8.51; -4.24†	-5.45	1.07	<0.001***	-7.54; -3.36
	URMA				**MRMA**			
TJC	-2.68	0.74	<0.001***	-4.12; -1.23	-2.56	0.71	<0.001***	-3.96; -1.17
SJC	-3.70	0.60	<0.001***	-4.87; -2.53	-3.50	0.57	<0.001***	-4.63; -2.38
Pain	-3.67	3.56	0.303	-10.63; 3.30	-3.45	3.62	0.341	-10.54; 3.65
Phy.Global	-0.33	0.27	0.233	-0.87; 0.21	-0.30	0.25	0.227	-0.79; 0.19
Pat.Global	-0.22	0.48	0.642	-1.15; 0.71	-0.24	0.41	0.567	-1.04; 0.57
Function	-0.05	2.97	0.988	-5.87; 5.78	-0.01	2.78	0.996	-5.47; 5.44
APR	-5.84	1.62	<0.001***	-9.03; -2.66	-5.57	1.69	0.001**	-8.89; -2.25

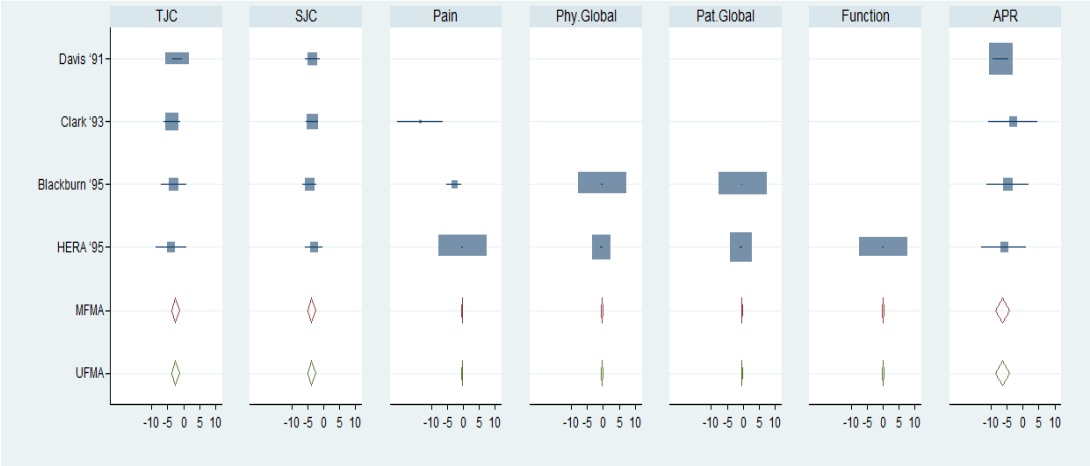
† Result from the original meta-analysis performed in the original systematic review

5.4.1.3 Summary of the differences between univariate and multivariate meta-analysis

In Figure 5.1, the forest plot for each outcome is reported. Each forest plot show for each study the MD and the pooled effect estimate obtained by applying UFMA and the MFMA models. In this figure, the pooled estimates obtained from the UFMA and MFMA models are considered. From Figure 5.1, it can be seen that for all the outcomes considered there is a benefit to be gained from treatment with Antimalarials. It is possible to see this benefit especially for the TJC, SJC and APR. Nevertheless, the 95% confidence interval associated with this estimate for the outcome function contains 0 and therefore it is not possible to reject with certainty the null hypothesis of no difference between the treatment with Antimalarials and the treatment with placebo (Table 5.2). When considering the MFMA approach, all outcomes had smaller SEs (increased precision) but appeared to be less efficacious than the equivalent UFMA. The highest difference in the estimates was for Pat.Global – from an estimate of the MD of -0.34 for UFMA to an estimate of the MD of -0.08 for MFMA. The statistical significance of the outcome Pat.Global was also overturned and was no longer significant, while the outcome pain was now only bordering on statistical significance.

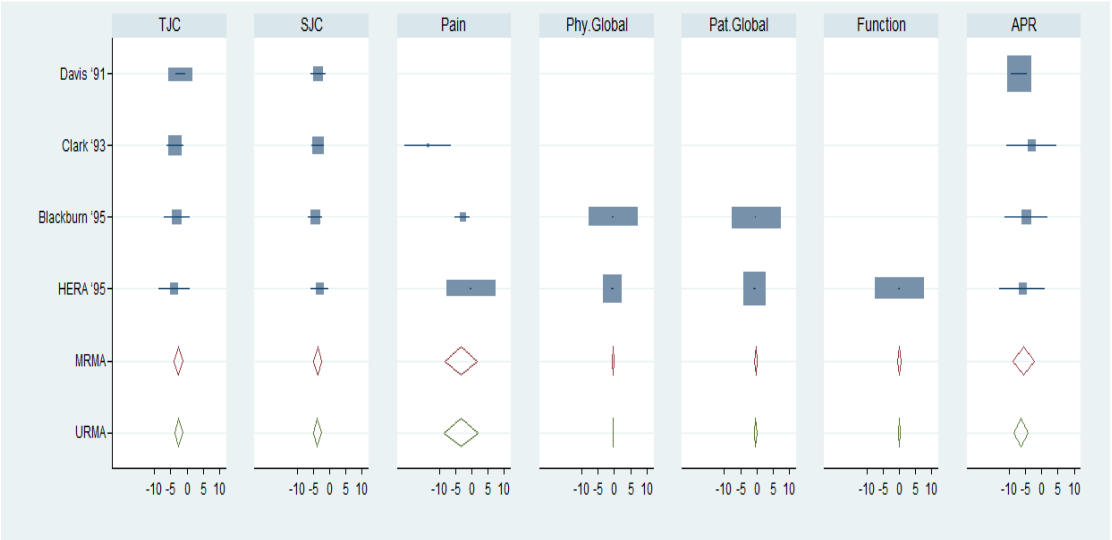
Applying MFMA, the MD estimate for function appeared to be (MFMA) also positive, denoting a favourable result (but not statistically significant) for the placebo treatment. It was interesting to note that these differences occurred in the outcomes where ORB was suspected, and the direction of the adjustment in applying the MVMA approach was consistent with the direction expected when ORB is present.

Figure 5.1 Antimalarials Forest Plots Fixed-Effects



In Figure 5.2 each forest plot reports for each study the MD and the pooled effect estimate obtained with the URMA and with the MRMA. As was seen also in Figure 5.1, in Figure 5.2 it is possible to observe that, for all the outcomes considered, the pooled estimates of the MD are lower than zero, indicating that there is a benefit in favour of the treatment with Antimalarials. It is possible to see this benefit especially for the APR. Nevertheless, the 95% confidence interval associated with the estimate of the MD for four outcomes – pain, the global outcomes (patient and physician) and function – contains 0 and therefore it is not possible to rule out the null hypothesis that the two treatments are not different.

Figure 5.2 Antimalarials Forest Plots Random-Effects



When the MRMA model was fitted, the I_R^2 value was 97%, indicating that, across all outcomes, the total variation in the meta-analysis is mainly due to between-study heterogeneity; this suggests that the REMA model might be the appropriate model to fit to these data. With the exception of the outcome pain, when comparing the MRMA approach and the URMA, all outcomes had smaller standard errors, meaning increased precision, but appeared to be less efficacious than the equivalent univariate analysis. The outcomes TJC, SJC count and APR were significant for URMA and are still significant for the MRMA approach.

5.4.2 Auranofin for treating rheumatoid arthritis

5.4.2.1 Summary of the outcome reporting bias assessment

The SR 'Auranofin for treating rheumatoid arthritis' conducted by Suarez-Almazor et al. [48] presents the results of UFMA.

The objective of this SR was to compare the short-term efficacy and toxicity of Auranofin for the treatment of RA by comparing auranofin with placebo. The SR aimed to synthesise data on all of the RA core outcomes, amongst a selection of other outcomes. This SR identified nine eligible studies, with a combined total of 539 randomised to auranofin and 510 to placebo. Two studies were excluded due to 'no OMERACT outcomes reported in this article'. Four studies were excluded because there was 'no placebo group'.

The results for the ORB assessment for the core outcomes are summarised in Table 5.3.

Table 5.3 ORB matrix for Auranofin systematic review

Study	TJC	SJC	Pain	Phy. Global	Pat. Global	Function	APR
Davies '82	✓	×(H)	✓	×(H)	×(H)	×(C)	×(A)
Palmer '82	✓	✓	×(C)	×(H)	✓	✓	✓
Prouse '82	✓	×(H)	✓	×(H)	×(H)	×(C)	✓
Lewis '84	×(F)	×(H)	×(E)	×(H)	×(H)	×(F)	✓
Johnsen '89	✓	✓	✓	×(D)	✓	×(FR)	✓
Ward '83	✓	✓	✓	✓	✓	×(FR)	✓
Wenger '83	✓	✓	✓	✓	×(H)	×(C)	✓
Bombardier '86	✓	✓	✓	✓	✓	✓	✓
Glennas '97	×(G)	✓	✓	×(H)	×(H)	✓	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Phy.: Physician; Pat.: Patient; APR: Acute Phase Reactant; FR: Fully reported
ORB Classification [9] A: Partial reporting, High Risk ORB; B: Partial reporting, No Risk ORB; C: Partial reporting, Low Risk ORB; D: None, High Risk ORB; E: None, High Risk ORB; F: None, Low Risk ORB; G: None, High Risk ORB; H: None, Low Risk ORB; I: NA, No Risk ORB.

Four studies were found to have low risk of ORB in one or more of the core outcomes according to the ORBIT classification system [28] (Palmer ’82; Prouse ’82; Ward ’83; Wenger ’83). The remaining four studies present a high risk of ORB for only one outcome (Davies ’82, APR; Lewis ’84, Pain; Johnsen ’89, Phy.Global; Glennnas ’97, TJC) and low risk of ORB classification for other outcomes.

5.4.2.2 Summary of the results from the original systematic review

In this particular review, the original study authors applied a FEMA to all seven core outcomes being considered. For pain (0%) and function (0%). Pat.Global (8%) and TJC (33%) the value of the heterogeneity index I^2 was low. While for remaining outcomes as SJC (68%), APR (68%) and Phy.Global (83%) the value of I^2 was high. For the outcome Pat.Global (51%) the value of the heterogeneity index was moderate. The pooled results for each outcome, as reported in the original review, can be found in Table 5.4 (UFMA). The use of Auranofin appears to be efficacious (the pooled estimates for the MD is negative) for the treatment of RA when considering all the core outcomes meta-analysed. The most efficacious effect of the treatment was calculated for the outcome APR (MD equal to -9.04) while the least efficacious effect of the treatment was estimated for the outcome function (MD equal to -0.13). Some estimates of the MD are statistically significant, this happened for TJC, pain, Pat.Global and Phy.Global and for APR. The remaining outcomes were characterised by non-significant estimates of the MD.

Table 5.4 Meta-analysis results for the Auranofin data set

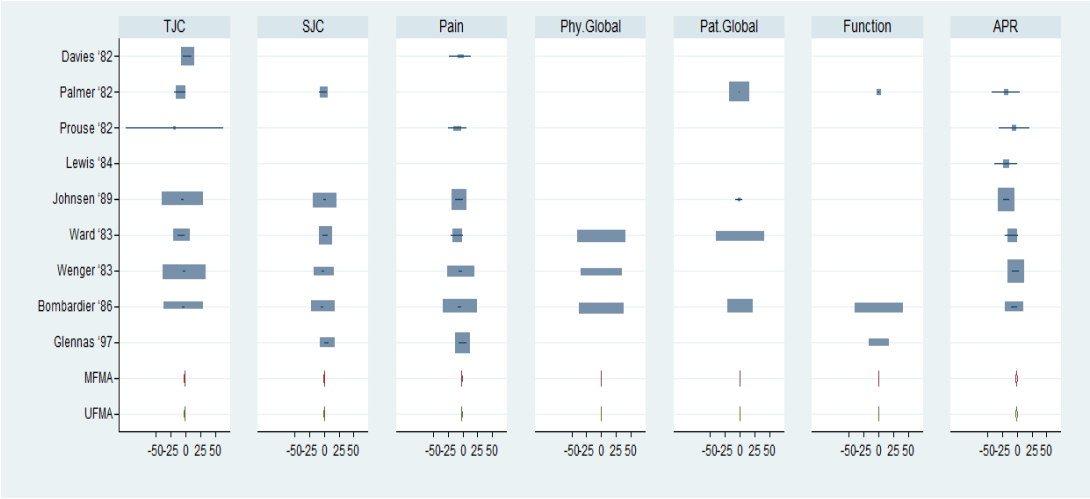
Outcome	Univariate meta-analysis				Multivariate meta-analysis			
	Mean difference	SE	P-value	95% CI	Mean difference	SE	P-value	95% CI
	UFMA				MFMA			
TJC	-3.76†	0.67	<0.001***	-5.06; -2.45†	-3.23	0.56	<0.001***	-4.32; -2.12
SJC	-0.29†	0.61	0.634	-1.49; 0.90†	-0.05	0.56	0.933	-1.14; 1.05
Pain	-4.68†	0.97	<0.001***	-6.59; -2.77†	-5.00	0.57	<0.001***	-6.11; -3.89
Phy.Global	-0.36†	0.08	<0.001***	-0.52; -0.21†	-0.28	0.06	<0.001***	-0.40; -0.15
Pat.Global	-0.41†	0.12	0.001**	-0.65; -0.17†	-0.24	0.06	<0.001***	-0.36; -0.13
Function	-0.13†	0.07	0.066	-0.27; 0.01†	-0.16	0.06	0.004**	-0.27; -0.05
APR	-9.04†	1.59	<0.001***	-12.16; -5.92†	-8.72	1.54	<0.001***	-11.74; -5.71
	URMA				MRMA			
TJC	-3.82	0.92	<0.001***	-5.63; -2.02	-3.62	0.97	<0.001***	-5.52; -1.73
SJC	0.15	1.18	0.900	-2.17; 2.46	0.41	1.17	0.726	-1.88; 2.69
Pain	-4.68	0.97	<0.001***	-6.59; -2.77	-4.44	1.07	<0.001***	-6.53; -2.35
Phy.Global	-0.39	0.21	0.058	-0.80; 0.01	-0.39	0.25	0.120	-0.89; 0.10
Pat.Global	-0.41	0.12	0.001**	-0.65; -0.17	-0.43	0.28	0.124	-0.98; 0.12
Function	-0.13	0.07	0.066	-0.27; 0.01	-0.21	0.11	0.054	-0.42; 0.00
APR	-9.79	3.25	0.003**	-16.16; -3.41	-10.26	3.22	0.001**	-16.57; -3.95

† Result from the original meta-analysis performed in the original systematic review

5.4.2.3 Summary of the differences between univariate and multivariate meta-analysis

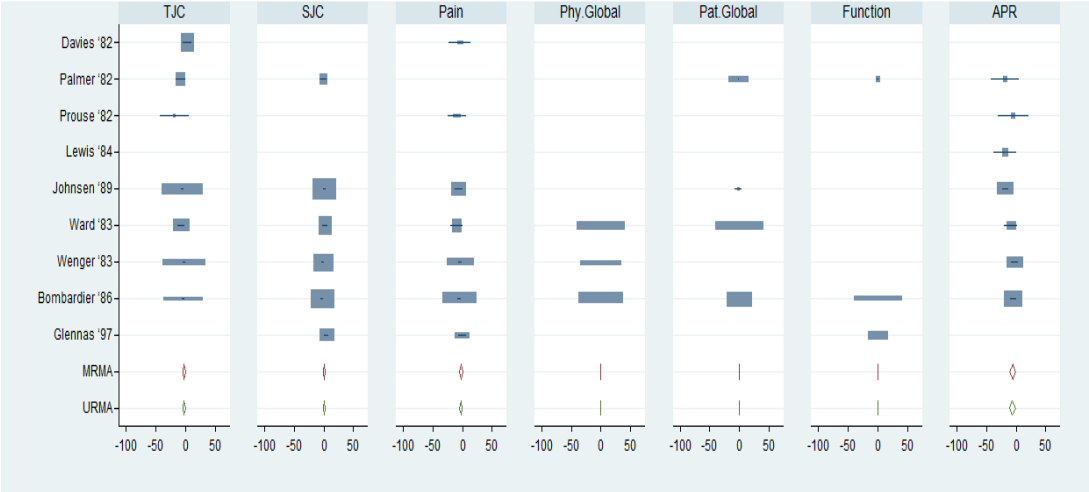
In Figure 5.3, the forest plots for the FEMA models applied are reported. When considering the MFMA approach, all outcomes had smaller standard errors (increased precision). With the exception of pain and function, all the other outcomes appeared to be less efficacious than the equivalent UFMA. The statistical significance of the outcome function was significant (0.004**) with the MFMA approach, while it was not significant (0.066) with the UFMA approach. This could be explained with the slight shift in the estimate of the MD away from the null from a value of -0.13 for UFMA to a result of -0.16 for MFMA. With the exception of the outcome pain, it is interesting to note that these differences occurred in the outcomes where ORB was suspected, and the direction of the adjustment in applying the MVMA approach was consistent with the direction expected when ORB is present.

Figure 5.3 Auranofin Forest Plots Fixed-Effects



In Figure 5.4, the forest plots for the REMA models applied are reported. When the MRMA model was fitted, the I_R^2 value was 83%, indicating that, across all outcomes, the total variation in the meta-analysis is mainly due to between-study heterogeneity; this suggests that the REMA model might be the appropriate model to fit to these data. The global estimate of the MD calculated suggested a benefit in the direction of the treatment with Auranofin for TJC, pain, Phy.Global, Pat.Global, function and APR, as for both URMA and MRMA it was less than zero. However, for TJC and pain, the estimates calculated with the MRMA model show that the treatment with Auranofin is less effective, while for Pat.Global, function and APR the MD estimates obtained with the MRMA model indicate that the treatment with Auranofin is more effective. The global MD for SJC was positive, indicating no effect of the treatment with Auranofin. When comparing the MRMA approach and the URMA, the outcomes TJC, pain, Phy.Global, Pat.Global and function had higher standard error applying the MRMA approach than the URMA model. On the other side, the outcomes SJC and APR had smaller standard error using the MRMA approach than the URMA approach. Pat.Global was significant for the URMA model (p-value equal to 0.001**) and switched to be not significant for the MRMA (p-value equal to 0.124) approach.

Figure 5.4 Auranofin Forest Plots Random-Effects



5.4.3 Leflunomide for the treatment of rheumatoid arthritis

5.4.3.1 Summary of the outcome reporting bias assessment

The SR 'Leflunomide for treating rheumatoid arthritis was conducted by Osiri et al. [55]. This SR presents the results of UFMA.

The objective of this SR was to assess the efficacy and toxicity of Leflunomide (monotherapy or combined with another DMARD) compared to placebo or other DMARDs in the treatment of RA. The SR aimed to synthesise data on all of the RA core outcomes, amongst a selection of other outcomes. This SR identified 33 eligible studies, while 24 studies were excluded. The five RCTs considered are all trials that test the same difference in the effectiveness of treatment with leflunomide against the treatment methotrexate. Al these trials considered measured and analysed the difference between these two treatments after 6 months of follow-up.

For the application of the UVMA and MVMA five studies were extracted from the 33 eligible RCT considered for the assessment of ORB and presented in Chapter 2. For the five studies considered (Strand '99, Bao '00, Lao '01, Shuai '02, Wislowska '07) when TJC, SJC, pain and APR were analysed, 390 subjects were randomised to the treatment with Leflunomide

and 373 were randomised to the alternative treatment with MTX. When Phy.Global and Pat.Global were analysed 360 subjects were randomised to the treatment with Leflunomide and 343 were randomised to the alternative treatment with MTX. Finally for the outcome function 104 subjects were randomised to the treatment with Leflunomide and 104 were randomised to the alternative treatment with MTX.

Three of the studies considered measures and reported on the seven core outcomes set (COS) (Strand ’99, Lao ’01, Shuai ‘02). Two studies were found to have high risk of ORB in one or more of the core outcomes according to the ORBIT classification system [28] (Bao ’00 – see Function; Wislowska ’07 – see the global measurements patient and Phy.Global) and no risk of ORB classification for the other outcomes.

Table 5.5 ORB matrix for Leflunomide systematic review

Study	TJC	SJC	Pain	Phy. Global	Pat. Global	Function	APR
Strand ‘99	✓	✓	✓	✓	✓	×(FR)	✓
Bao ‘00	✓	✓	✓	✓	✓	×(E)	✓
Lao ‘01	✓	✓	✓	✓	✓	✓	✓
Shuai ‘02	✓	✓	✓	✓	✓	✓	✓
Wislowska ‘07	✓	✓	✓	×(G)	×(E)	✓	✓

TJC: Tender Joint Count; SJC: Swollen Joint Count; Phy.: Physician; Pat.: Patient; APR: Acute Phase Reactant; FR: Fully reported
ORB Classification [9] A: Partial reporting, High Risk ORB; B: Partial reporting, No Risk ORB; C: Partial reporting, Low Risk ORB; D: None, High Risk ORB; E: None, High Risk ORB; F: None, Low Risk ORB; G: None, High Risk ORB; H: None, Low Risk ORB; I: NA, No Risk ORB.

Observing the table reported (Table 5.5) in this section, it is possible to notice that TJC, SJC, pain and APR were FR in all studies of this SR.

5.4.3.2 Summary of the results from the original systematic review

In this review, the original study authors applied a FEMA to all seven core outcomes being considered. For TJC (0%), SJC (0%), Phy.Global (0%), Pat.Global (0%) and APR (20%) the value of the heterogeneity index I^2 was low. For the outcome function (53%) the value of I^2 was moderate. The pooled results for each outcome, as reported in the original review,

can be found in Table 5.6 (UFMA). The use of Leflunomide appears to be efficacious for the treatment for RA when considering TJC, pain, Phy.Global, Pat.Global and function (slightly -0.01), while the use of alternative treatment appears to be efficacious when examining SJC (0.14) and APR (1.54). For Phy.Global and Pat.Global, the estimates of the MD are statistically significant, with a p-value equal to 0.005 (Phy.Global) and 0.001 (Pat.Global), while for all the other outcomes the MDs are not significant.

Table 5.6 Meta-analysis results for the Leflunomide data set

Outcome	Univariate meta-analysis				Multivariate meta-analysis			
	Mean difference	SE	P-value	95% CI	Mean difference	SE	P-value	95% CI
	UFMA				MFMA			
TJC	-0.64†	0.41	0.112	-1.44; 0.15†	-0.13	0.35	0.715	-0.82; 0.56
SJC	0.14†	0.31	0.654	-0.47; 0.75†	0.42	0.29	0.148	-0.15; 1.00
Pain	-0.32†	0.24	0.180	-0.78; 0.15†	-0.08	0.14	0.565	-0.35; 0.19
Phy.Global	-0.48†	0.17	0.005**	-0.82; -0.15†	-0.30	0.13	0.002**	-0.56; -0.05
Pat.Global	-0.60†	0.18	0.001**	-0.95; -0.26†	-0.33	0.11	0.021*	-0.54; -0.12
Function	-0.01†	0.05	0.889	-0.11; 0.09†	0.04	0.04	0.275	-0.03; 0.11
APR	1.54†	1.50	0.647	-1.41; 4.48†	1.59	1.45	0.270	-1.25; 4.44
	URMA				MRMA			
TJC	-0.64	0.41	0.116	-1.44; 0.16	-0.10	0.37	0.788	-0.83; 0.63
SJC	0.11	0.34	0.735	-0.55; 0.78	0.40	0.32	0.202	-0.22; 1.02
Pain	-0.34	0.25	0.181	-0.83; 0.16	-0.05	0.16	0.759	-0.37; 0.27
Phy.Global	-0.49	0.18	0.001**	-0.83; -0.14	-0.31	0.14	0.005**	-0.57; -0.04
Pat.Global	-0.62	0.19	0.006**	-1.00; -0.24	-0.33	0.12	0.024*	-0.55; -0.10
Function	-0.01	0.07	0.925	-0.15; 0.14	0.03	0.06	0.587	-0.09; 0.15
APR	0.52	2.53	0.838	-4.45; 5.49	1.35	2.53	0.594	-3.62; 6.32

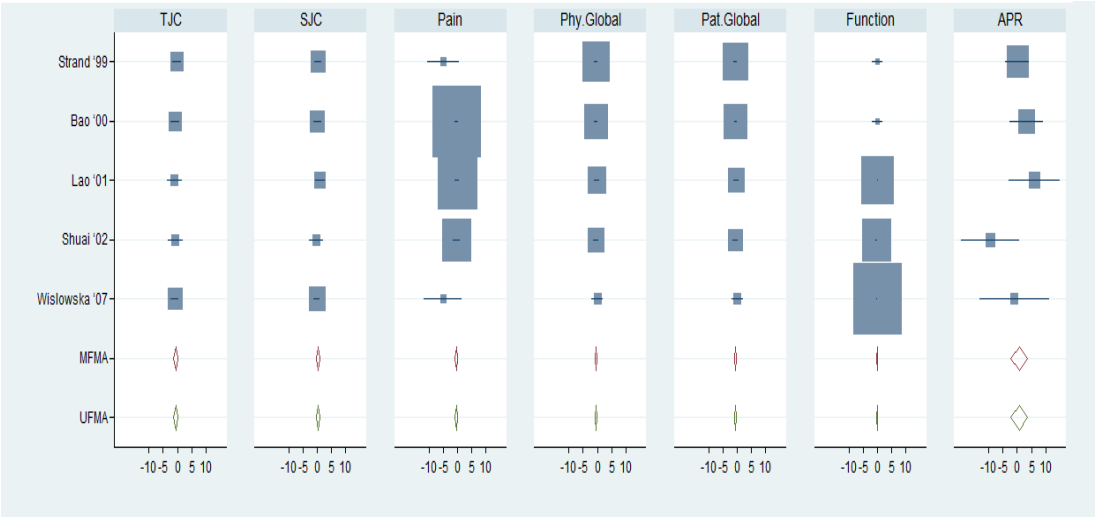
† Result from the original meta-analysis performed in the original systematic review

5.4.3.3 Summary of the differences between univariate and multivariate meta-analysis

In Figure 5.5, the forest plots for the FEMA models applied are reported. When considering the MFMA approach, for TJC, pain, Phy.Global and Pat.Global the efficacy of the treatment with Leflunomide decreases. For SJC and for APR, where the benefit is in favour of alternative treatment with MTX, this result is more noticeable with the MFMA approach than with the UFMA. For the outcome function, there is a change in the direction of effectiveness, which is in favour of the treatment with Leflunomide (-0.01) for the UFMA approach while it is in favour for the treatment with MTX (0.04) for the MFMA approach. For all outcomes considered in the analysis, a reduction in the standard error (higher precision) may be

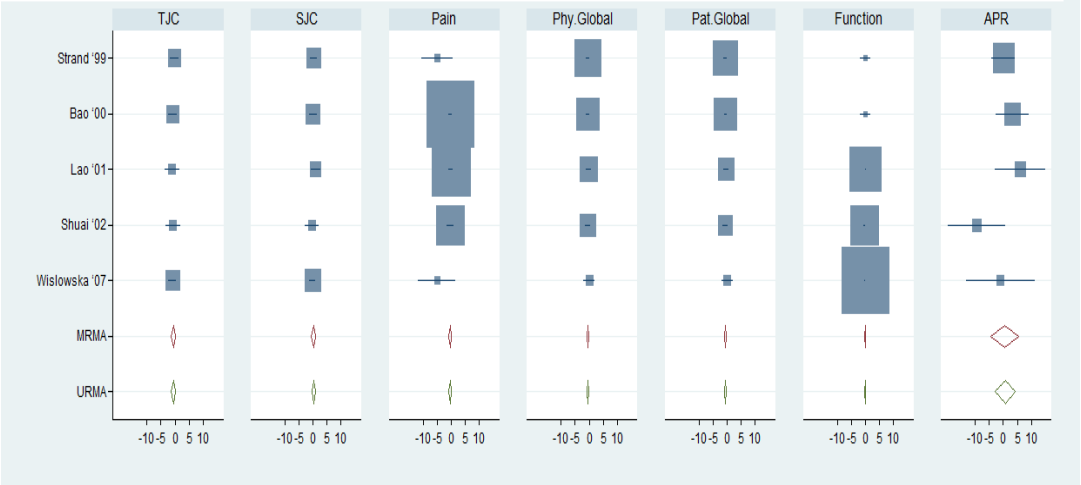
noticed when comparing from the MFMA to the UFMA model approach. The outcomes Phy.Global and Pat.Global are the only two that have significant estimates for both the studied models.

Figure 5.5 Leflunomide Forest Plots Fixed-Effects



In Figure 5.6, the forest plots for the REMA models applied are reported. When the MRMA model was fitted, the I_R^2 value was 34%, indicating that, across all outcomes, the total variation in the meta-analysis is mainly due to between-study heterogeneity; this suggests that the REMA model might be the appropriate model to fit to these data. As was described before, also for the comparison between MRMA and URMA, for TJC, pain, Phy.Global and Pat.Global the treatment with Leflunomide appears to be less efficacious. For SJC and APR, where the efficacy was in the opposite direction in favour of the alternative treatment, the alternative treatment seems to be more efficacious with the MRMA approach than with the URMA approach. For function, the estimate of the MD was negative for the URMA (-0.01) while it was positive for the MRMA (0.03). With the exception of APR, all the estimates appear to be more precise (as the Standard Errors decrease) using the MRMA approach rather than the URMA approach. The MD for the outcomes Phy.Global and Pat.Global are statistically significant for both the URMA and MRMA approaches.

Figure 5.6 Leflunomide Forest Plots Random-Effects



5.4.4 Methotrexate for treating rheumatoid arthritis

5.4.4.1 Summary of the outcome reporting bias assessment

The SR 'Methotrexate for treating rheumatoid arthritis' was conducted by Suarez-Almazor et al [52]. This SR presents the results of UFMA.

The objective of this SR was to evaluate the short term efficacy and toxicity of MTX for the treatment of RA. This SR identified five eligible studies with a combined total of 113 randomised to MTX and 106 to placebo.

Observing the table reported (Table 5.7) in this section, it can be seen that TJC and function were FR in all (100%) studies of this SR. In addition, the other outcomes considered showed a high percentage of fully reporting, as 80% of the studies of this SR were reporting these outcomes.

Table 5.7 ORB matrix for Methotrexate systematic review

Study	TJC	SJC	Pain	Phy. Global	Pat. Global	Function	APR
Andersen '85	✓	✓	✓	✓	✓	✓	✓
Weinblatt '85	✓	✓	×(H)	✓	✓	✓	✓
Williams '85	✓	✓	✓	✓	✓	✓	×(C)
Furst '90	✓	✓	✓	✓	✓	✓	✓
Pinheiro '93	✓	×(G)	✓	×(H)	×(H)	✓	✓

TJC: Tender Joint Count; SJC: Swollen Joint Count; Phy.: Physician; Pat.: Patient; APR: Acute Phase Reactant; FR: Fully reported
ORB Classification [9] A: Partial reporting, High Risk ORB; B: Partial reporting, No Risk ORB; C: Partial reporting, Low Risk ORB; D: None, High Risk ORB; E: None, High Risk ORB; F: None, Low Risk ORB; G: None, High Risk ORB; H: None, Low Risk ORB; I: NA, No Risk ORB.

Two of the studies considered measured and reported on the seven core outcomes set (Andersen '85 and Furst '90). Two studies were found to have low risk of ORB in one of the core outcomes according to the ORBIT classification system [28] (Weinblatt '85 see Pain; Williams '85 see APR) and no risk of ORB classification for the other outcomes. One study (Pinheiro '93) was found to have high risk of ORB in one of the core outcomes according to the ORBIT classification system [28] (SJC) and low risk of ORB classification for two outcomes (patient and physician's global) and no risk for the other outcomes.

5.4.4.2 Summary of the results from the original systematic review

In this review, the original study authors applied a FEMA to all seven core outcomes being considered. For the outcomes Function (95%), Pat.Global (94%), Phy.Global (93%), pain (915) and TJC (68%) the heterogeneity index I^2 was high. For the outcome APR (56%) the index I^2 was moderate and finally for the outcome SJC (31%), the value of I^2 was low. The pooled results for each outcome, as reported in the original review, can be found in Table 5.8 (UFMA). The use of methotrexate appears efficacious for all the outcomes considered in this systematic review. Some outcomes have a substantial value for the MD, such as TJC (-17.85), APR (-8.95) and pain (-3.0). With the exception of 'APR', the estimates of the MD are statistically significant (p-value is < 0.001***).

Table 5.8 Meta-analysis results for the Methotrexate data set

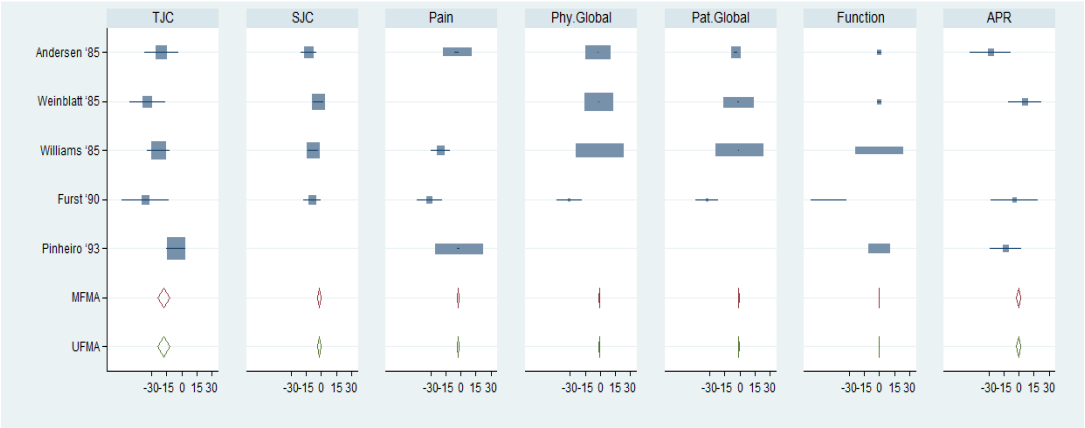
Outcome	Univariate meta-analysis				Multivariate meta-analysis			
	Mean difference	SE	P-value	95% CI	Mean difference	SE	P-value	95% CI
	UFMA				MFMA			
TJC	-17.85†	3.12	<0.001***	-23.97; -11.73†	-5.33	2.70	0.048*	-10.62; -0.05
SJC	-7.31†	1.60	<0.001***	-10.44; -4.18†	-1.38	1.35	0.310	-4.03; 1.28
Pain	-3.00†	0.54	<0.001***	-4.07; -1.93†	-1.25	0.37	0.001**	-1.97; -0.53
Phy.Global	-1.05†	0.13	<0.001***	-1.31; -0.80†	-0.39	0.10	0.040*	-0.60; -0.20
Pat.Global	-0.92†	0.15	<0.001***	-1.20; -0.63†	-0.16	0.08	<0.001***	-0.30; -0.01
Function	-0.48†	0.05	<0.001***	-0.58; -0.38†	-0.39	0.04	<0.001***	-0.47; -0.31
APR	-8.95†	4.70	0.057	-18.17; 0.27†	1.80	4.51	0.690	-7.03; 10.63
	URMA				MRMA			
TJC	-22.63	7.00	0.001**	-36.35; -8.90	-19.19	6.21	0.002**	-31.36; -7.01
SJC	-8.74	2.37	<0.001***	-13.38; -4.09	-6.88	2.07	0.001**	-10.94; -2.82
Pain	-15.20	6.22	0.015*	-27.40; -3.00	-15.46	6.90	0.025*	-28.99; -1.94
Phy.Global	-5.01	5.73	0.374	-17.07; 5.37	-5.85	5.91	0.310	-16.59; 6.57
Pat.Global	-6.53	7.35	0.397	-20.93; 7.86	-7.59	7.47	0.307	-22.24; 7.06
Function	-12.61	10.35	0.223	-32.89; 7.66	-9.52	10.17	0.349	-29.45; 10.41
APR	-7.41	7.27	0.308	-21.65; 6.83	-6.72	6.75	0.320	-19.95; 6.52

† Result from the original meta-analysis performed in the original systematic review

5.4.4.3 Summary of the differences between univariate and multivariate meta-analysis

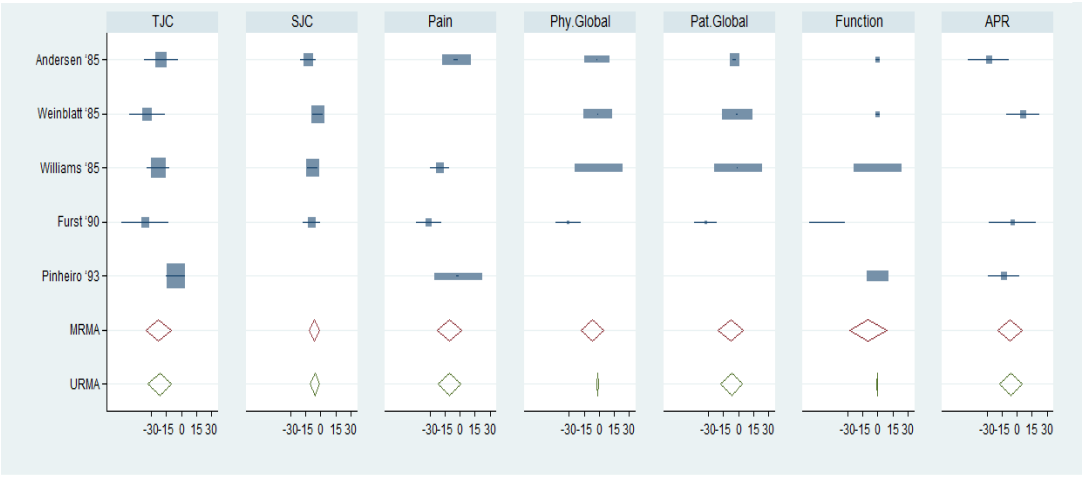
In Figure 5.7, the forest plots for the FEMA models applied are reported. When considering the MFMA approach, it is possible to notice how for TJC, SJC, pain, Pat.Global, Phy.Global and function there is a change in the estimates of the MD and the effectiveness of treatment with methotrexate decreases. The outcome APR showed an estimate of the MD (-8.95) that suggested efficacy of the treatment with Methotrexate with the UFMA approach, while it showed an estimate of the MD (1.80) that suggested efficacy of the alternative treatment with the MFMA approach. For all the outcomes, the estimate of the standard error calculated with the MFMA approach is lower than the standard error calculated with the UFMA, and therefore the estimates of the MD calculated for the MFMA approach are more precise than the estimates of the MD calculated with UFMA. With the exception of SJC, all the estimates that were statistically significant with the UFMA approach are also significant with the MFMA approach. The estimates of the MDs calculated for the outcome APR are not significant, neither with the UFMA nor with the MFMA.

Figure 5.7 Methotrexate Forest Plots Fixed-Effects



In Figure 5.8, the forest plots for the REMA models applied are reported. When we fitted the MRMA, the value of I_R^2 was equal to 99.5%, indicating that, across all outcomes, the total variation in the meta-analysis is mainly due to between-study heterogeneity; this suggests that the REMA model might be appropriate to fit to these data. From the comparison between the MRMA and URMA models, the treatment with MTX appears to be less efficacious with the MRMA model than with the URMA for the outcomes TJC, SJC, function and APR, while the treatment with Methotrexate seems to be more efficacious with the MRMA model than with URMA for the outcomes pain, Phy.Global and Pat.Global. The outcomes TJC, SCJ, function and APR have more precise estimates of the MD with the MRMA than with the URMA model, while the outcomes pain, Phy.Global and Pat.Global have more precise estimates of the MD with the URMA than with the MRMA. The outcomes TJC, SJC and pain appear to have significant estimates of the MD for both the models considered, URMA and MRMA.

Figure 5.8 Methotrexate Forest Plots Fixed-Effects



5.4.5 Sulfasalazine for treating rheumatoid arthritis

5.4.5.1 Summary of the outcome reporting bias assessment

The SR 'Sulfasalazine for treating rheumatoid arthritis' was conducted by Suarez-Almazor et al [54]. This SR presents the results of UFMA.

The aim of this SR was to estimate the short-term efficacy and toxicity of Sulfasalazine for the treatment of RA. For the analysis to obtain the results included in this section, attention was focused on the comparison Sulfasalazine vs. Placebo. This SR considered six eligible trials, including 468 patients. These patients were randomised to Sulfasalazine (243) and to Placebo (225).

As can be seen in the following table, this SR considered six studies. Six outcomes were considered and analysed as none of the studies examined functional outcomes with comprehensive functional scales, and therefore this outcome could not be adequately assessed in the meta-analysis. Observing the table reported in this section (Table 5.9), it is possible to notice that TJC was FR in 67% of the studies in this SR, while SJC, pain, and APR were FR in half of the included studies (50%). Pat.Global and Phy.Global were FR in 33% of the studies. All the studies considered had at least one of the six core outcomes set at risk of ORB according to the ORBIT classification system [28]. All the studies considered

in this SR were found to have at least one study with a high risk of ORB classification. For one study (Pullar ’83) considered, all the COS were found to have a high risk of ORB classification. For another study (Skosey ’88) considered, two of the COS were found to have a high risk of ORB classification (TJC and SJC), and the remaining outcomes were found to have a low risk of ORB. Farr ’95 was found to have three outcomes (SJC, Pat.Global and Phy.Global) with a high risk of ORB classification. Ebringer ’92, the global measurements (Pat.Global and Phy.Global) were found to have a high risk of ORB classification. Two studies (Williams ’88 and Hannonen ’93) were found to have a high risk of ORB only for one outcome each, respectively Pain and APR.

Table 5.9 ORB matrix for Sulfasalazine systematic review

Study	TJC	SJC	Pain	Phy. Global	Pat. Global	APR
Williams ‘88	✓	✓	×(G)	✓	✓	✓
Ebringer ‘92	✓	✓	✓	×(D)	×(D)	✓
Hannonen ‘93	✓	✓	✓	✓	✓	×(D)
Farr ‘95	✓	×(G)	✓	×(D)	×(D)	✓
Pullar ‘83	×(E)	×(G)	×(E)	×(G)	×(G)	×(E)
Skosey ‘88	×(A)	×(A)	×(C)	×(C)	×(C)	×(C)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Phy.: Physician; Pat.: Patient; APR: Acute Phase Reactant; FR: Fully reported
ORB Classification A: Partial reporting, High Risk ORB; B: Partial reporting, No Risk ORB; C: Partial reporting, Low Risk ORB; D: None, High Risk ORB; E: None, High Risk ORB; F: None, Low Risk ORB; G: None, High Risk ORB; H: None, Low Risk ORB; I: NA, No Risk ORB.

5.4.5.2 Summary of the results from the original systematic review

In this review, the original study authors applied a FEMA to all seven core outcomes being considered. For the outcomes Pat.Global (78%), TJC (66%) and APR (62%) the value heterogeneity index I^2 was high. For the outcome Phy.Global (46%) the value of the index I^2 was moderate and finally for the outcomes SJC (18%) and pain (0%), the value of I^2 was low. The pooled results for each outcome, as reported in the original review, can be found in Table 5.10 (UFMA). All the outcomes considered and analysed in this systematic review show an efficacious result for the treatment with Sulfasalazine as the estimated MDs are all lower than zero. In particular, this efficacy is noticeable for APR (-17.58) and pain (-

8.71). Treatment with Sulfasalazine has moderate efficacy for TJC (-2.45) and SJC (-2.38), and finally the efficacy of the treatment with Sulfasalazine appears to be slight for the outcomes Phy.Global (-0.16), and Pat.Global (-0.23). With the exception of Phy.Global and Pat.Global, the other outcomes present estimates of the MD that are statistically significant.

Table 5.10 Meta-analysis results for the Sulfasalazine data set

Outcome	Univariate meta-analysis				Multivariate meta-analysis			
	Mean difference	SE	P-value	95% CI	Mean difference	SE	P-value	95% CI
	UFMA				MFMA			
TJC	-2.45†	0.87	0.005**	-4.15; -0.74†	-0.95	0.78	0.223	-2.48; 0.58
SJC	-2.38†	0.69	0.001**	-3.73; -1.03†	-1.45	0.62	0.019*	-2.67; -0.24
Pain	-8.71†	3.11	0.005**	-14.80; -2.62†	-1.52	2.56	0.552	-6.53; 3.49
Phy.Global	-0.16†	0.11	0.140	-0.37; 0.05†	-0.03	0.09	0.747	-0.20; 0.15
Pat.Global	-0.23†	0.12	0.052	-0.46; 0.00†	-0.15	0.10	0.117	-0.35; 0.04
APR	-17.58†	2.22	<0.001***	-21.93; -13.23†	-15.03	2.04	<0.001***	-19.04; -11.03
	URMA				MRMA			
TJC	-4.29	2.41	0.075	-9.01; 0.43	-4.40	2.76	0.111	-9.81; 1.02
SJC	-2.72	0.96	0.004**	-4.59; -0.85	-2.20	0.79	0.006**	-3.76; -0.64
Pain	-8.84	3.88	0.023*	-16.44; -1.24	-14.20	10.31	0.169	-34.41; 6.01
Phy.Global	-0.19	0.14	0.169	-0.46; 0.08	-0.20	0.17	0.219	-0.53; 0.12
Pat.Global	-0.31	0.18	0.089	-0.67; 0.05	-0.33	0.20	0.098	-0.73; 0.06
APR	-15.07	5.74	0.009**	-26.32; -3.83	-17.10	2.69	<0.001***	-22.37; -11.83

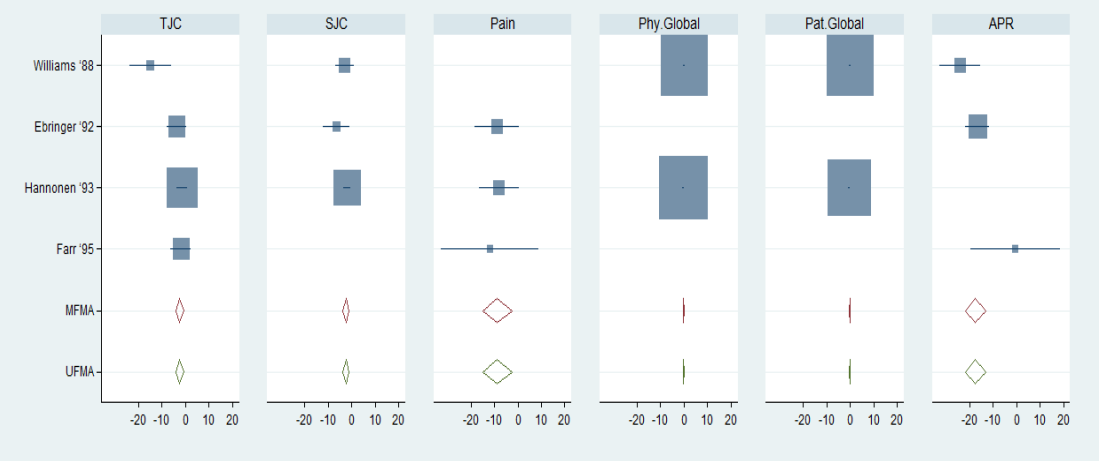
† Result from the original meta-analysis performed in the original systematic review

5.4.5.3 Summary of the differences between univariate and multivariate meta-analysis

In Figure 5.9, the forest plots for the FEMA models applied are reported. When considering the MFMA approach, it is possible to notice how for all the outcomes considered in this meta-analysis the estimates of the MD decrease, suggesting that the treatment with Sulfasalazine appears to be less efficacious with an MFMA approach than with a UFMA approach. All the outcomes taken into account in this meta-analysis show how the estimate of the MD appears to be more precise with an MFMA approach than with a UFMA one, as the Standard Errors calculated by the MFMA model are lower than the Standard Errors calculated by the UFMA. For two outcomes, TJC and pain, the estimates of the MD were significant for the UFMA approach as they were presenting a p-value equal to 0.005. For the other outcomes considered there were no changes, and therefore the estimates of the MDs

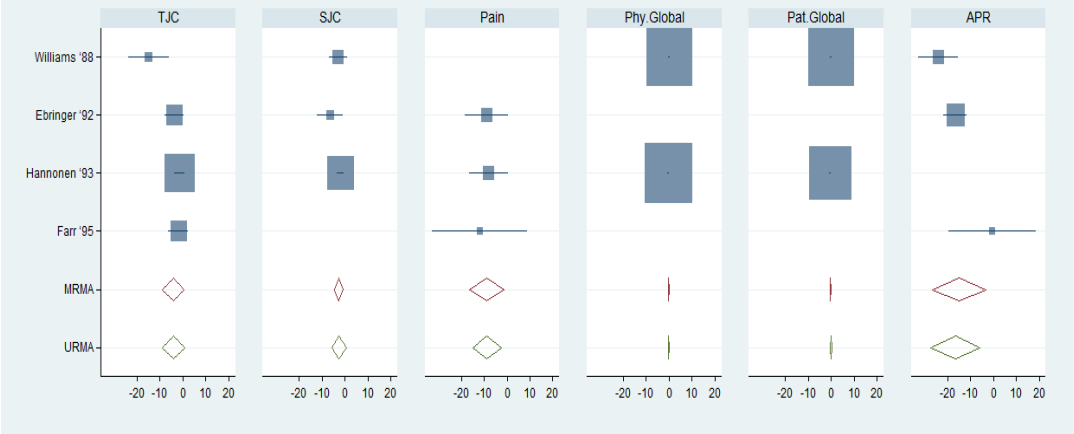
were significant for both UFMA and MFMA for SJC and APR, while they were not significant for both UFMA and MFMA for Phy.Global and Pat.Global.

Figure 5.9 Sulfasalazine Forest Plots Fixed-Effects



In Figure 5.10, the forest plots for the REMA models applied are reported. When the MRMA was fitted, the value was 49.1%, indicating that, across all outcomes, 49.1% of the total variation in the meta-analysis is due to between-study heterogeneity. The observation of estimates calculated by the MRMA model suggests that, with the exception of the outcome SJC, there is an increase in the estimated efficacy of the treatment with Sulfasalazine. The estimates of the MD, switching from URMA to MRMA model, become more precise for SJC and APR, while for the other outcomes they become less precise as the Standard Errors calculated increase. In particular, for outcome pain, which was seen to be characterised by high suspicion of ORB, the Standard Error calculated increases by 62% from an SE equal to 3.88 calculated by URMA to an SE equal to 10.31 calculated by the MRMA model. Observing the results for the outcome pain is interesting also regarding the significance of the estimates, as it is possible to notice how the MD estimated by the URMA model was significant (p-value equal to 0.023) while it was not significant when the MRMA model was applied (p-value equal to 0.169). For other outcomes considered, the statistical significance of the estimates does not change, and therefore the MDs calculated are still significant for the outcomes SJC and APR, while they are still not significant for the outcomes TJC, Phy.Global and Pat.Global.

Figure 5.10 Sulfasalazine Forest Plots Random-Effects



5.4.6 Adalimumab for treating rheumatoid arthritis

5.4.6.1 Summary of the outcome reporting bias assessment

The SR 'Adalimumab for treating rheumatoid arthritis' was conducted by Navarro-Sarabia et al [57]. This SR presented the results of URMA.

The aim of this review was to assess the efficacy and safety of adalimumab in the treatment of RA. Therefore six studies with 2381 patients were included in this review. Two comparisons were done: A. adalimumab subcutaneously (sc) + methotrexate (or DMARDs) versus placebo sc + MTX (or DMARDs). B. Adalimumab sc in monotherapy versus placebo sc.

To run the model UVMA and MVMA for this example, attention was focused on the comparison between treatment with Adalimumab sc + MTX (or DMARDs) and Placebo sc + MTX (or DMARDs) and the results reported in the original SR have been considered for the comparison between Adalimumab sc in monotherapy and placebo. Therefore, Van De Putte '03 (all seven COS) and Van De Putte '04 (six outcomes: TJC, SJC, pain, Pat.Global, Phy.Global and Function) were considered FR. for the application of univariate and MVMA models. As can be seen in the following table (Table 5.11), this SR considered six studies.

Observing the table reported in this section, it is possible to notice that among the outcome that were considered six of them (TJ, SJ, pain, Pat.Global, Phy.Global and function) were characterised by a 67% percentage of fully reporting. The fully reporting percentage for the APR outcome was 50%. Two studies (Weinblatt ’03 and Keystone ’04) measured and reported on the seven COS. All the remaining studies considered in this SR were found to have outcomes with a low risk of ORB classification. We discussed previously Van de Putte ’03 and Van de Putte ’04. Furthermore, Furst ’03 (for all outcomes ‘F’ classification) and Rau ’04 (for all outcomes ‘C’ classification) have to be considered.

Table 5.11 ORB matrix for Adalimumab systematic review

Study	TJC	SJC	Pain	Phy. Global	Pat. Global	Function	APR
Weinblatt ‘03	✓	✓	✓	✓	✓	✓	✓
Keystone ‘04	✓	✓	✓	✓	✓	✓	✓
Furst ‘03	×(F)	×(F)	×(F)	×(F)	×(F)	×(F)	×(F)
Van de Putte ‘03	×(F.R)	×(F.R)	×(F.R)	×(F.R)	×(F.R)	×(F.R)	×(F.R)
Van de Putte ‘04	×(F.R)	×(F.R)	×(F.R)	×(F.R)	×(F.R)	×(F.R)	×(C)
Rau ‘04	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Phy.: Physician; Pat.: Patient; APR: Acute Phase Reactant; FR: Fully reported
ORB Classification [9] A: Partial reporting, High Risk ORB; B: Partial reporting, No Risk ORB; C: Partial reporting, Low Risk ORB; D: None, High Risk ORB; E: None, High Risk ORB; F: None, Low Risk ORB; G: None, High Risk ORB; H: None, Low Risk ORB; I: NA, No Risk ORB.

5.4.6.2 Summary of the results from the original systematic review

In this review, the original study authors applied a FEMA to all seven core outcomes being considered. For the outcomes APR (77%) and Phy.Global (63%) the value of heterogeneity index I^2 was high. For the remaining outcomes SJC (19%), TJC (0%), pain (0%), Pat.Global (0%) and function (0%), the value of the index I^2 was low. The pooled results for each outcome, as reported in the original review, can be found in Table 5.12 (URMA).

Table 5.12 Meta-analysis results for the Adalimumab data set

Outcome	Univariate meta-analysis				Multivariate meta-analysis			
	Mean difference	SE	P-value	95% CI	Mean difference	SE	P-value	95% CI
	UFMA				MFMA			
TJC	-7.50	2.06	<0.001***	-11.53; -3.47	-7.51	2.05	<0.001***	-11.52; -3.50
SJC	-6.28	1.49	<0.001***	-9.21; -3.36	-6.28	1.49	<0.001***	-9.20; -3.37
Pain	-16.02	2.77	<0.001***	-21.44; -10.60	-16.05	2.76	<0.001***	-21.46; -10.64
Phy.Global	-20.00	4.54	<0.001***	-28.91; -11.10	-20.01	4.53	<0.001***	-28.89; -11.14
Pat.Global	-18.29	3.77	<0.001***	-25.68; -10.90	-18.31	3.76	<0.001***	-25.68; -10.95
Function	-0.33	0.06	<0.001***	-0.45; -0.22	-0.33	0.06	<0.001***	-0.45; -0.22
APR	-1.23	0.48	0.010*	-2.17; -0.29	-1.23	0.48	0.010*	-2.16; -0.29
	URMA				MRMA			
TJC	-6.68	1.20	<0.001***	-9.02; -4.33	-6.84	1.19	<0.001***	-9.18; -4.50
SJC	-5.75	0.88	<0.001***	-7.475; -4.020	-5.78	0.88	<0.001***	-7.51; -4.06
Pain	-15.79	2.28	<0.001***	-20.26; -11.32	-16.19	2.28	<0.001***	-20.65; -11.73
Phy.Global	-19.42	2.40	<0.001***	-27.19; -11.65	-18.21	2.06	<0.001***	-22.24; -14.19
Pat.Global	-17.01	2.40	<0.001***	-22.05; -13.98	-17.35	2.39	<0.001***	-22.04; -12.66
Function	-0.33	0.05	<0.001***	-0.42; -0.23	-0.33	0.05	<0.001***	-0.42; -0.24
APR	-1.21	0.20	<0.001***	-2.09; -0.33	-1.07	0.20	<0.001***	-1.46; -0.68

† Result from the original meta-analysis performed in the original systematic review

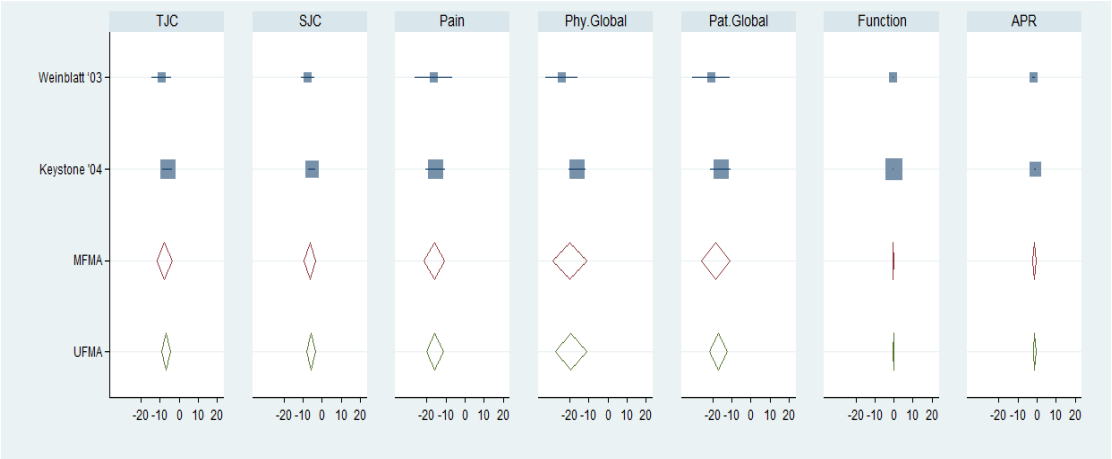
All the outcomes considered were showing a benefit from the combination treatment of Adalimumab and MTX (or DMARDs) as the estimate of the MD was lower than zero. In particular, the combination therapy of Adalimumab and MTX (or DMARDs) is highly effective with regard to pain, Phy.Global and Pat.Global. The therapy assessed is mildly effective with regard to TJC and SJC, and finally it is slightly effective with regard to APR and function.

5.4.6.3 Summary of the differences between univariate and multivariate meta-analysis

Figure 5.11 shows the forest plots for this systematic review considering all the outcomes assessed and only the two studies for which all the outcomes were FR (Weinblatt ’03 and Keystone ‘04). Therefore, Furst ’03, Van de Putte ’03, Van de Putte ’04 and Rau ’04 are not reported and not considered as they present missing values for all the outcomes. When the MFMA model was applied, the estimates of the MD slightly increased for the outcomes TJC, pain, Phy.Global and Pat.Global, suggesting higher efficacy for the combination therapy of Adalimumab and MTX (or DMARDs). For all the remaining outcomes, there was no change in the estimates of the MD comparing UFMA and MFMA. The estimates of the MD for the outcomes TJC, pain, Phy.Global and Pat.Global are also characterised by an increase in

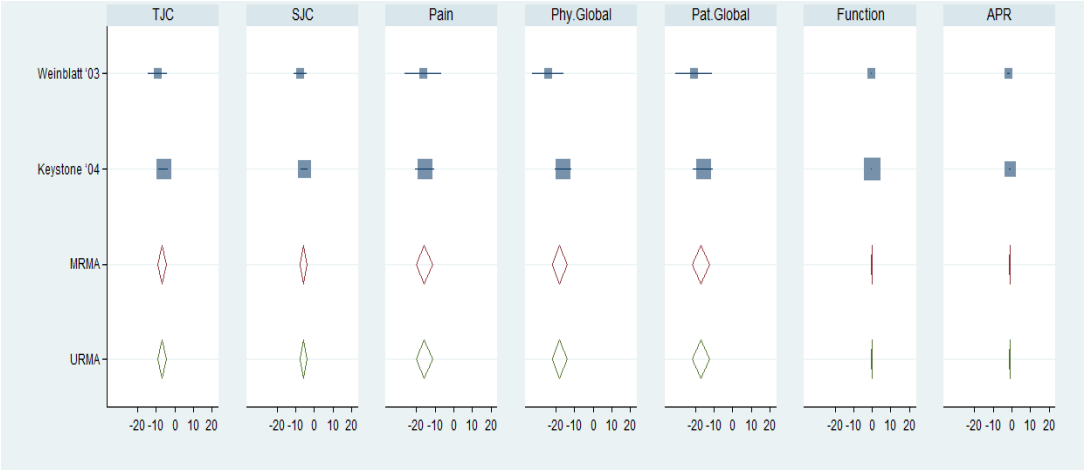
precision as the Standard Errors calculated for the estimates of these outcomes slightly decreased. All the estimates of the MD were extremely significant, indicating clear support for combination therapy with Adalimumab and MTX (or DMARDs). This happened for both the UFMA and MFMA approaches.

Figure 5.11 Adalimumab Forest Plots Fixed-Effects



Additionally, Figure 5.12 shows the forest plots for this systematic review considering all the outcomes assessed and only the two studies for which all the outcomes were FR. The observation of estimates calculated by the MRMA model suggests that the combination treatment of Adalimumab and MTX (or DMARDs) appears to be more efficacious for the outcomes TJC, SJC, pain and Pat.Global, while the studied treatment appears to be less efficacious for the outcomes Phy.Global and APR. There is no change in the estimate of the MD comparing URMA and MRMA for the outcome function. Comparing URMA and MRMA, the estimates of the MD appear to be more precise for TJC, Phy.Global and Pat.Global as the Standard Errors calculated for these outcomes decrease. For all the other outcomes considered, there is no change in the estimates of the precision as the Standard Errors are equal across the URMA and MRMA models fitted. All the estimates of the MD were extremely significant, indicating clear support for combination therapy with Adalimumab and MTX (or DMARDs). This happened for both the URMA and MRMA approaches.

Figure 5.12 Adalimumab Forest Plots Random-Effects



5.4.7 Abatacept for treating rheumatoid arthritis

5.4.7.1 Summary of the outcome reporting bias assessment

The SR 'Abatacept for treating rheumatoid arthritis' was conducted by Maxwell and Singh [58]. This SR presented the results of UFMA.

The aim of this SR was to assess the efficacy and safety of Abatacept in reducing disease activity, pain, and improving function in people with RA. In this SR seven trials with 2908 patients were included. The patients were randomised to treatment with Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic (1444) and to the treatment with placebo + DMARDs/biologic (757). Some outcomes were originally reported by the reviewers as Pat.Global (Weinblatt '07) and Function (Genovese '05, Weinblatt '07 and Schiff '08); the comparison was not of interest.

Therefore, for these cases the notation in the table will be ×(FR), which means that the outcome was measured and reported but in this analysis it has not been considered (no risk of ORB). As can be observed in the following table (Table 5.13), this SR considered seven studies.

Table 5.13 ORB matrix for Abatacept systematic review

Study	TJC	SJC	Pain	Phy. Global	Pat. Global	Function
Weinblatt '06	×(E)	×(E)	✓	✓	✓	✓
Kremer '03	✓	✓	✓	✓	✓	✓
Kremer '06	✓	✓	✓	✓	✓	✓
Moreland '02	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)
Genovese '05	×(E)	×(E)	×(E)	×(E)	×(E)	×(FR)
Weinblatt '07	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)
Schiff '08	× (E)	× (E)	×(E)	×(E)	×(E)	×(FR)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Phy.: Physician; Pat.: Patient; APR: Acute Phase Reactant; FR: Fully reported
ORB Classification [9] A: Partial reporting, High Risk ORB; B: Partial reporting, No Risk ORB; C: Partial reporting, Low Risk ORB; D: None, High Risk ORB; E: None, High Risk ORB; F: None, Low Risk ORB; G: None, High Risk ORB; H: None, Low Risk ORB; I: NA, No Risk ORB.

Six outcomes were considered and analysed as none of the studies examined APR and therefore this outcome could not be adequately assessed in the meta-analysis. Focusing attention on the percentage of reporting, it can be seen that Pat.Global and Phy.Global were FR in less than half (43%) of the studies included in this systematic review. In addition, TJ and SJ were FR in few (29%) of the included studies, while pain was FR in 57% of them, and finally function was FR in 86%. Observing the following table, it can be seen that two of the seven studies included in this analysis measured and reported on the seven COS (Kremer '03 and Kremer '06). Three studies were found to have high risk of ORB in one or more of the core outcomes according to the ORBIT classification [28]. For Genovese '05 and Schiff '08, five outcomes had high risk of ORB (83%) (TJC, SJC, Pain, Pat.Global, Phy.Global) and for Weinblatt there were two (33%) (TJC and SJC). Moreland '02 was found to have low risk of ORB ('×(C)' classification) in all the outcomes considered in this analysis. Weinblatt '07 was found to have no risk of ORB as the classification was FR for all the outcomes considered.

5.4.7.2 Summary of the results from the original systematic review

In this review, the original study authors applied a FEMA to six of the core outcomes being considered. The lower value of heterogeneity index I^2 was calculated for SJC (0%) while the higher value of heterogeneity was calculated for outcome TJC (76%). The pooled results for each outcome, as reported in the original review, can be found in Table 5.14 (UFMA).

Table 5.14 Meta-analysis results for the Abatacept data set

Outcome	Univariate meta-analysis				Multivariate meta-analysis			
	Mean difference	SE	P-value	95% CI	Mean difference	SE	P-value	95% CI
	UFMA				MFMA			
TJC	-7.28†	0.86	<0.001***	-8.96; -5.59†	-5.97	0.81	<0.001***	-7.55; -4.40
SJC	-4.78†	0.56	<0.001***	-5.87; -3.68†	-3.23	0.46	<0.001***	-4.14; -2.33
Pain	-12.44†	0.96	<0.001***	-14.32; -10.57†	-12.83	0.95	<0.001***	-14.69; -10.97
Phy.Global	-13.29†	0.85	<0.001***	-14.96; -11.61†	-13.11	0.85	<0.001***	-14.78; -11.45
Pat.Global	-12.46†	1.04	<0.001***	-14.49; -10.42†	-12.49	1.03	<0.001***	-14.51; -10.47
Function	-0.39†	0.05	<0.001***	-0.48; -0.30†	-0.39	0.05	<0.001***	-0.48; -0.30
	URMA				MRMA			
TJC	-7.31	1.84	<0.001***	-10.92; -3.70	-7.18	1.94	<0.001***	-10.97; -3.38
SJC	-5.11	0.94	<0.001***	-6.94; -3.27	-4.98	1.07	<0.001***	-7.08; -2.88
Pain	-15.13	3.79	<0.001***	-22.56; -7.70	-15.10	3.78	<0.001***	-22.51; -7.70
Phy.Global	-15.75	5.11	0.002**	-25.76; -5.74	-15.66	5.13	0.002**	-25.71; -5.61
Pat.Global	-14.37	2.96	<0.001***	-20.17; -8.57	-14.38	2.89	<0.001***	-20.05; -8.72
Function	-0.49	0.16	0.003**	-0.81; -0.17	-0.49	0.17	0.004**	-0.82; -0.15

† Result from the original meta-analysis performed in the original systematic review

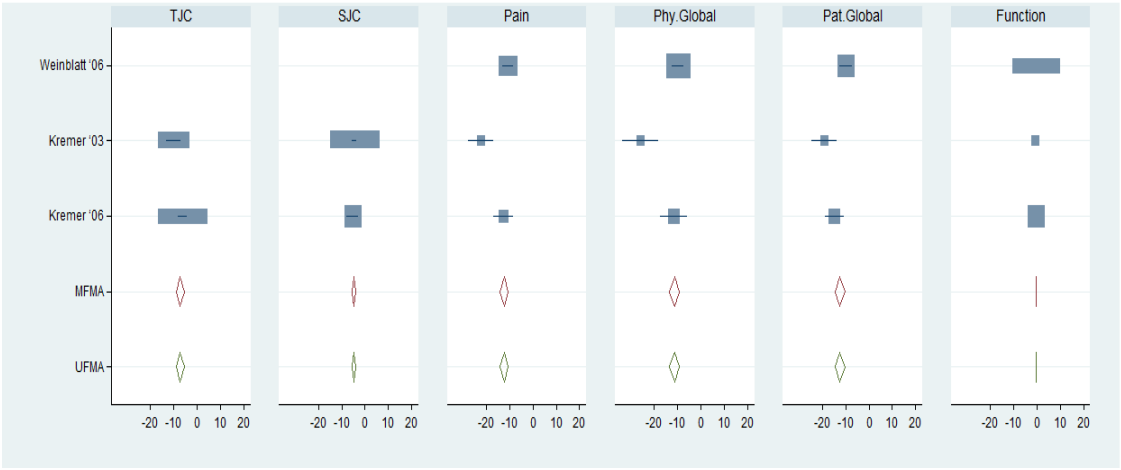
All the outcomes considered are showing a benefit from the combination treatment of Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic as the estimate of the MD is lower than zero. In particular, the combination therapy of Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic is highly effective with regard to pain, Phy.Global and Pat.Global. The therapy assessed is mildly effective with regard to TJC and SJC, and finally it is slightly effective with regard to function.

5.4.7.3 Summary of the differences between univariate and multivariate meta-analysis

Figure 5.13 shows the forest plots for this systematic review considering all the outcomes assessed and excluding the four studies for which the outcomes were missing (Moreland

’02, Genovese ’05, Weinblatt ’07, Schiff ’08). When the MFMA model was applied, the estimates of the MD decreased for the outcomes TJC, SJC and Phy.Global, suggesting a lower efficacy of the combination therapy of Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic. For pain and Pat.Global, the estimates of the MD increased, indicating a higher efficacy of the combination therapy of Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic. For function there is no change in the estimate of the MD comparing the UFMA and MFMA models. With the exception of Phy.Global and function, the estimates of the MD are characterised by an increase in precision as the Standard Errors calculated for the estimates of these outcomes decreased. All the estimates of the MD were extremely significant, indicating clear support for combination therapy with Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic. This happened for both the UFMA and MFMA approaches.

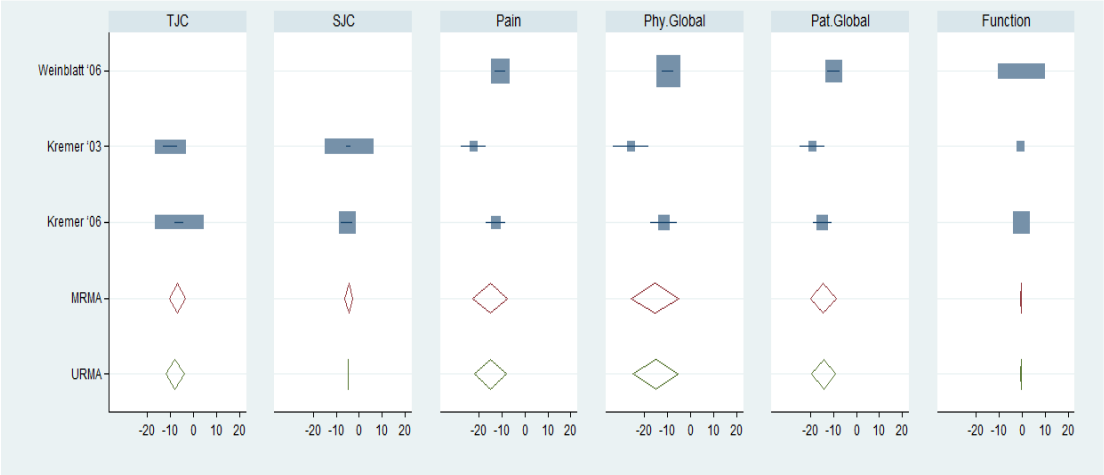
Figure 5.13 Abatacept Forest Plots Fixed-Effects



Additionally, Figure 5.14 shows the forest plots for this systematic review considering all the outcomes assessed and excluding the four studies for which the outcomes were missing. When we fitted the MRMA, the value was 98.5%, indicating that, across all outcomes, the total variation in the meta-analysis is mainly due to between-study heterogeneity; this suggests that the REMA model might be appropriate to fit to these data. When the MRMA model was applied, the estimates of the MD decreased for the outcomes TJC, SJC, pain and Phy.Global, suggesting a lower efficacy of the combination therapy of Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic. For the outcomes TJC, SJC, Phy.Global and function, the estimates of the MD calculated with the MRMA model appear to be less

precise, as the SEs calculated using the MRMA model are higher than the SEs obtained fitting the URMA. For the outcomes pain and Pat.Global, the estimates of the MD calculated with the MRMA model appear to be more precise. All the estimates of the MD were extremely significant, indicating clear support for combination therapy with Abatacept (2 mg/kg and 10 mg/kg) + DMARDs/biologic. This happened for both the URMA and MRMA approaches.

Figure 5.14 Abatacept Forest Plots Random-Effects



5.5 Discussion

This is the first study to consider an assessment of ORB against a well-established core set of outcomes. The OMERACT COS (sometimes referred to as the International League of Associations for Rheumatology core set of outcomes) was ratified for use in clinical trials for rheumatoid arthritis but is also endorsed by the CMSG [30]. Although the uptake of the measurement of the COS for rheumatoid arthritis trials has been shown to be increasing [91], the reporting of the outcomes for many of these trials remains insufficient, meaning that many meta-analyses are unable to include data from all relevant studies. The MVMA approach offers one such sensitivity analysis to reduce the potential impact of ORB when there are missing trial data for many review outcomes.

From the results obtained in this study, the MVMA and the UVMA have been applied to seven (33%) systematic reviews previously assessed for the presence of ORB. From the observation of the results from the UVMA and MVMA, it is possible to see that for three (43%) of the systematic reviews analysed there was a change in the direction of the effect, comparing results of the UVMA and MVMA models. This happens not for all the outcomes considered in the assessment. This change happened for both FEMA and REMA approaches for one SR, the SR that studied Leflunomide for the treatment of RA. For this SR, the outcome function was presenting a negative estimate (-0.01) of the MD when UVMA was applied, while it was showing a positive estimate of the MD when MVMA was applied (equal to 0.04 for MFMA and 0.03 for MRMA). The change in the direction of the estimates happened only for the FEMA approach for the SR that studied Antimalarials for the treatment of RA. For this SR, the outcome function was presenting a negative estimate (-0.06) of the MD when UFMA was applied, while it was showing a positive estimate of the MD (0.10) when MFMA was applied. Furthermore, for the SR that studied Methotrexate for the treatment of RA the change in the direction of the estimates happened only for the FEMA approach. For this SR, the outcome APR was presenting a negative estimate (-8.95) of the MD when UFMA was applied, while it was showing a positive estimate of the MD (1.80) when MFMA was applied. From the observation of the results from the UVMA and MVMA, it is possible to see that for four (57%) of the systematic reviews analysed there was a change in the significance of the estimates, comparing results of the UVMA and MVMA models. This happens not for all the outcomes considered in the assessment. With the exception of one case, this change was from a significance of the estimates of the UVMA to no significance of the estimates of the estimates form the MVMA. For two SRs, this change occurred for both FEMA and REMA approaches. These two SRs were Sulfasalazine and Auranofin. Observing the results from the Auranofin systematic review [48] for the treatment of rheumatoid arthritis, it is possible to see that the change in the significance happened for the outcome function, which was presenting a non-significant estimate with the UFMA while it was showing a significant estimate with the MFMA. Furthermore, the outcome Pat.Global was showing a significant estimate for URMA and a non-significant estimate with the MRMA. Commenting on the results from the Sulfasalazine systematic review for the treatment of

rheumatoid arthritis, it is possible to determine that the change in the significance happened for outcome TJC and pain that were presenting a significant estimate with the UFMA while were showing a non-significant estimate with the MFMA. Furthermore, the outcome pain was showing a significant estimate for URMA and a non-significant estimate with the MRMA. Additionally, the Antimalarials and the Methotrexate systematic reviews for the treatment of rheumatoid arthritis were showing a significant estimate for UFMA and a non-significant result for the MFMA model.

According to the work of ORB assessment conducted on the 47 outcomes considered in the analysed SRs for this application, 20 outcomes (43%) were found to be suspected of high risk of bias. Among these outcomes, the direction of the adjustment was in the expected direction (toward the null) for 17 (85%) outcomes. For the outcome function, as was discussed before, there is a change in the direction of the estimate from a negative estimate to a positive estimate of the MD. However, some limitations to this approach need to be kept in mind. The first is that the ORB assessment is subjective. The second is that the MVMA does not directly model the ORB mechanism; it simply utilises the extra information from the w/s correlation to adjust the effect estimates (and this does not necessarily control the direction of the change). The third is that MVMA does not distinguish between high/low ORB.

With the exception of the Adalimumab Systematic review, all the remaining original reviews used FEMA methods for all outcomes in all examples. Of particular interest was the observation of the difference in the results between URMA and MRMA models; it is possible to say that there are changes in the treatment estimates between URMA and MRMA. For example, for the Auranofin systematic review [48] it was seen that the estimates moved towards the null for TJC and pain. Furthermore, results showed a shift in treatment effect estimates that was in the opposite direction for the global measurements (physician and patient), function and APR. Finally, observing the results for the outcome function, the shift in direction to a more positive result of the estimate does not suggest ORB.

Finally, it has been seen from the results obtained that the MVMA model (fixed and random effects) is not always simple to apply to these examples, as there were some issues and some limitations in the application of this model. In particular, it needs to be recalled that these models are fitted using AD that does not give information about the w/s correlation between the outcomes considered in the model. This issue needs to be solved before fitting MVMA. As previously highlighted in the methods section, other researchers have defined different methods, such as calculating the w/s correlation directly in each study based on the availability of IPD, approximating the w/s correlation using biological reasoning or expert opinion, and finally using the Pearson correlation method when it is impossible to obtain w/s correlation from IPD or from expert opinion [74]. When obtaining the results for this chapter, the w/s correlations were calculated based on Pearson correlations calculated at patient level from a previous study conducted by Professor Dr. George Wells (unpublished data). The Pearson correlations calculated at patient level have been used in the Wei and Higgins formula (5.3) described in Section 5.3.5 of this chapter to obtain the desired w/s correlations to be used in the MVMA models.

Chapter 6 - Discussion

This final chapter provides a summary of the thesis and a discussion of its key findings and limitations. This chapter also aims to explain what the work adds to the research field of evidence-based synthesis. Recommendations for further work are also provided.

6.1 Overview of the thesis and how it adds to evidence-based synthesis

It was stated in Chapter 1 of this thesis that systematic reviews represent the gold standard method for reviewing research literature [1]. Many systematic reviews contain meta-analysis, which is the statistical component that is used to synthesise the results of independent studies [7]. Meta-analysis and systematic reviews (SR) can make important contributions to medical research by showing that there is evidence to support treatments not widely used, or that evidence is lacking to support treatments that are in wide use [7].

The reliability of systematic review conclusions is largely based on finding data on a complete and unbiased set of studies [7]. Missing data from systematic reviews can heavily impact on results and can affect systematic reviews in two ways. Publication bias, where a study is not published on the basis of its results, can lead to bias in the analysis of a particular outcome in a review, especially if the decision not to submit or publish the study is related to the results for that outcome [28]. A second form of missingness comes from published studies, for example, when only a subset of the originally recorded outcome variables in a trial are selectively reported in the publication based on their results. When the decision not to report on certain outcomes is driven by the significance and/or direction of the effect size (e.g., non-significant outcomes are reported only as p-value <0.05 or are suppressed altogether), then this is referred to as outcome reporting bias (ORB) or selective outcome reporting bias [28]. This thesis has considered the problem of ORB. The impact of ORB has previously been shown to overturn findings and to overestimate treatment effects [28], although in this thesis we consider alternative impact methods to those previously

used. Missing participant data [33] may also be considered as a third source of missingness within the systematic review process if IPD-MA are performed, which has been examined in Chapter 4.

6.1.1 Assessment of ORB in a cohort of systematic reviews for the treatment of rheumatoid arthritis

In Chapter 2, the process for assessing ORB in systematic reviews was described, which includes the most up-to-date methods and tools available for researchers to use. The methods were then applied to a cohort of Cochrane systematic reviews considering pharmacological interventions for rheumatoid arthritis. Outcome reporting bias was assessed against a well-established set of eight core outcomes, which is commonly known as the World Health Organisation (WHO) and International League of Associations for Rheumatology (ILAR) core set of outcomes [29]. This core outcome set has not only been approved for use in clinical trials for rheumatoid arthritis but is also endorsed by the Cochrane Musculoskeletal Group. The use of COS is important for this application of meta-analysis because it reduces waste in research which in turn can reduce the impact of ORB [29].

Although the development of the COS for rheumatoid arthritis has shown that the measurement of the outcomes is increasing [29], the reporting of the outcomes for many of these trials remains insufficient, meaning that many meta-analyses are unable to include data from all relevant studies. Chapter 2 contains the results and the discussion of the work on ORB assessment conducted in this thesis when applying the ORBIT classification system [28].

The ORB assessment was applied in 21 included reviews, and all contained missing data on at least one of the eight outcomes. This assessment has found that ORB was highly suspected in 275 (23%) of the 1205 evaluable outcomes from 167 assessable trials. At the trial level, missing or incomplete reporting of outcome data for each core outcome ranged

from 41% (tender joint count) to 59% (Phy.Global). For 49.4% of the evaluable outcomes in our study, the set of core outcomes was either partially reported or not reported (A to I classification). For 23%, at least one core outcome was classified under high suspicion for ORB (A, D, E, or G), whereas for 19.6% it was clear the outcomes were measured and analysed (A, B, C, D), but the reporting of the outcomes meant that the data could not be included in a meta-analysis. Similar to the study reported by Dwan et al. [92], all the reviews considered in this study included at least one study that contained missing data in relation to at least one of the core outcomes. The results of the work on ORB assessment were published in 2014 by Frosi et al. [36].

6.1.2 Methods for studying the outcome reporting bias issue in systematic reviews

In addressing the issue of ORB, there are two different solutions. The first can be defined as a non-statistical solution and consists of contacting trialists to obtain missing outcome data. If it is not feasible to obtain the missing data from the trialist, there are solutions that could be used, for example, statistical solutions such the maximum bias bound [28], a multivariate meta-analysis (MVMA) approach [16], or a model-based correction method [70]. The assumption of the maximum bias bound approach is that larger studies (with a small standard error) are more likely to be published than smaller studies (with a large standard error) [28]. The model-based correction method proposed by Copas et al. [71] assumes that, if a paper in the area of interest does not report sufficiently well (or not at all) the particular outcome of interest, then either (i) the outcome was measured but failed to show a significant treatment effect or (ii) the outcome was simply not measured [71]. Omitting such a study from the meta-analysis will lead to a bias in the overall treatment effect in case (i) but not in case (ii). If (i) is known to be true, the bias will be in the direction away from the null, resulting in overestimation of the treatment effect and exaggeration of significance [71]. Previous studies showed that the results from the MVMA approach were encouraging and it was found that the ‘borrowing of strength’ through the use of correlation between outcomes can reduce the impact of ORB in a meta-analysis [16]. Therefore, the purpose of this thesis was to apply MVMA to reduce the impact of outcome reporting bias in randomised controlled

trials for rheumatoid arthritis. Therefore, multivariate meta-analysis was firstly applied to some scenarios concerning missing data in simulated SRs with individual participant data, and then it was applied to some SRs for the assessment of pharmacological treatment for rheumatoid arthritis.

6.1.3 Simulation of missing data scenarios in individual participant data and application of two-stage individual participant data meta-analysis

In Chapter 3 and Chapter 4, an empirically-based individual participant data (IPD) simulation study based on the rheumatoid arthritis data was conducted using a fixed-effects meta-analysis approach. The potential benefits of multivariate meta-analysis were examined over standard univariate analyses for a variety of different ORB missing data mechanisms.

In Chapter 4, further simulations were considered to see if any of the benefits of multivariate meta-analysis extended to missing participant data within IPD.

Previous research [16] has examined the benefits of bivariate fixed-effects meta-analysis (BFMA) over univariate fixed-effects meta-analysis (UFMA) for estimating pooled (treatment) effects for multiple outcomes in both complete and missing data scenarios. This work contained certain limitations and uncovered a number of questions. One of the limitations was that in their simulation work the researchers generated aggregate data (AD) instead of considering individual participant data (IPD). The basic assumption when simulating AD is that the within-study correlations are the same in each meta-analysis dataset which may not be realistic [16].

The IPD meta-analysis simulation study that was performed demonstrated that the MFMA was preferential compared to the UFMA across the majority of performance criteria and scenarios considered. The benefit was particularly noticeable as the ORB missing data mechanism became more severe and the patient-level correlation between the outcomes increased.

The results from the simulation study were encouraging and, considering the mean square error (MSE) as a trade-off between preferred bias and improved precision, then the multivariate meta-analysis approach was preferred over the two different univariate specifications irrespective of an MCAR or MAR missing data mechanism and two different levels of missing data (20% and 40%). The benefit of MFMA when looking at %BoS was also clear to see, particularly as the amount of missing data and the patient-level correlation between outcomes increases. More BoS occurred in outcomes when the missingness was MCAR than when it was MAR.

6.1.4 Application of univariate meta-analysis and multivariate meta-analysis to specific examples of systematic reviews assessed

In the final experimental chapter, Chapter 5, the multivariate meta-analysis approach was applied to the motivating examples of rheumatoid arthritis. From the observation of the results from the UVMA and MVMA, it is possible to see that for three (43%) of the seven SRs analysed there was a change in the direction of the effect, comparing results of the UVMA and MVMA models. This happens not for all the outcomes considered in the assessment.

6.1.4.1 Fixed-effects meta-analysis

According to the assessment of these SRs, the analysed results showed that 20 (43%) of the total outcomes considered (47: five SRs considered seven outcomes and 2 SRs considered six outcomes therefore 47 come from the calculation 7*5+6*2) were found to be at high suspicion of ORB. From the comparison of MFMA and UFMA it is possible to see that, for 17 (85%) out of the 20 outcomes suspected to be at high risk of ORB, there is a change in the estimates towards the null. Furthermore, the results showed that the precision of the estimates increased for 19 (95%) of the outcomes suspected of having a high risk of ORB.

6.1.4.2 Random-effects meta-analysis

For 17 (85%) out of the 20 outcomes suspected to be at high risk of ORB there is no change in the significance of the estimates. From the comparison of MRMA and URMA it is possible to see that, for 13 (65%) out of the 20 outcomes suspected to be at high risk of ORB, there is a change in the estimates towards the null. Furthermore, the results showed that the precision of the estimates increased for 10 (50%) of the outcomes suspected of high risk of ORB. For 19 (95%) out of the 20 outcomes suspected to be at high risk of ORB there is no change in the significance of the estimates.

6.2 Use of multivariate meta-analysis to limit issues of outcome reporting bias

This thesis was mainly dedicated to the investigation of the potential benefits that MVMA may offer to address the problem of ORB (and missing participant data) from primary studies that plagues many systematic review meta-analyses.

MVMA has many potential advantages over separate univariate meta-analyses [16]. This thesis has shown that MVMA allows outcome data to be synthesised jointly in the same model; thus, the pooled estimates for all outcomes are obtained simultaneously, whilst accounting for their correlation. The BoS across different outcomes given the correlations allows for the inclusion of studies with missing outcome data, which would have otherwise been lost using standard UVMA methods. This was a particular strength of the MVMA method in the presence of ORB, and therefore may lead to more representative inferences and potential bias reduction than the standard univariate approach. These benefits may also equally translate across to missing participant data if more complex IPD meta-analyses are being performed.

6.3 Limitations of multivariate meta-analyses

Through the simulation studies, MVMA has been shown to be beneficial compared to standard UVMA, particularly in the presence of ORB across a number of different performance criteria; however, there are some limitations attached to this approach.

A possible limitation of this work is that this study focused on all Cochrane reviews (up to and including September 2012 issue) that had considered only pharmacological interventions for the treatment of RA [39]. Despite the scope of the RA COS, there was also a suggestion that the COS were being more frequently measured in RA trials that the RA COS was not specifically designed for, for example, non-pharmacological treatments [39]. However, the reporting of the full RA COS for these intervention types in general remained low [39]. For example another study [39] found that for many of the non-pharmacological trials the participants included in the study had RA (hence the RA COS was relevant), but the nature of the intervention meant that many of the primary investigators were not rheumatologists [39].

A methodological limitation of this work of thesis is that univariate analyses are thought to be easier to understand and more transparent to the average meta-analyst (who may not have a statistical background). MVMA methods for obtaining w/s correlations required for the analysis but rarely reported or made available in primary studies may also provide a suitable barrier to successful and wide implementation of this approach. The availability of IPD means that these w/s correlations can be computed directly (as was performed in the simulations study (Chapters 3 and 4)), but these complex review methods are rarely carried out due to the availability of IPD, and therefore such correlations may need to be estimated as was the case for the applications in Chapter 5. A full review of how to estimate w/s correlations for use in MVMA can be found in Wei and Higgins' paper [74]. Another limitation is that the differences in the statistical properties between the univariate and MVMA approach are sometimes only marginal, especially when there are no missing data.

Perhaps the most notable limitation of the MVMA approach as a method of the reduction of the impact or ORB on systematic review conclusions is that the direction of the adjustment cannot be controlled by the method (it is a consequence of utilising the w/s correlation).

Finally it is possible to say that typically, ORB overestimates treatment effects when it is present (i.e. due to the removal of non-significant results) [28], and therefore a suitable adjustment method, under this assumption, would shift the adjusted estimates back towards the null, of no treatment effect difference. While this trend was observed in the simulation studies, where the true missing data mechanism was known, and the performance measures were averaged over many simulations, it was not always observed for a single practical application when high risk of ORB was suspected in an outcome. The ORB classifications themselves for judging high and low risk of bias can also be subjective, and may contribute to this underlying issue, although the sensitivity and specificity of judging high/low risk of bias using this classification system has previously been found to be high [28]. Notably, a more recent ORB adjustment approach developed by Copas [71] specifically directly models the ORB missing data mechanism.

6.4 Further work

This thesis has made a useful contribution to the understanding of MVMA methods as a method of ORB adjustment, but many issues remain unaddressed. While the REMA, MVMA model was introduced in the applications chapter, for simplicity this approach was not investigated as part of a larger simulation study. In the presence of ORB, a FEMA may not be a realistic assumption in practice as ORB may hide the true extent of underlying heterogeneity, suggesting that a REMA approach could be the default approach in the presence of ORB. Considering the REMA, MVMA methods need to be investigated further, to see if the same benefits observed in the fixed-effects simulations are still apparent.

Furthermore, there needs to be a comparison of the various ORB adjustment approaches (inclusive of MVMA) available to see which method is preferred and which one can be more

easily applied. In addition, it would be beneficial to understand more clearly, when MVMA should be considered over UVMA.

Assessing ORB for harmful outcomes in reviews has also been considered [69], and an investigation into the use of MVMA as an adjustment approach should be considered in this situation when the reason for missing outcome data may not be based on statistical significance.

6.5 Conclusions

MVMA is not the solution to all ORB-related problems within review meta-analyses, but, informed by this thesis, it can unquestionably be seen as a route to address missing outcome data. The BoS, reduction in bias and increase in precision can all be seen as promising, despite some of the aforementioned challenges with the approach [73].

As well as developing statistical methodology such as MVMA to adjust for ORB, promotion of preventative methods should continue.. It has been stated that this is an important step for controlling the impact of ORB [29]. The strategies proposed to reduce the impact of ORB included trial registration, outcome data banks and online journals with greater space to include more information [13].

APPENDIX A

ASSESSMENT OF A COHORT OF SYSTEMATIC REVIEWS OF TREATMENTS FOR REHUMATOID ARTHRITIS

List of systematic reviews assessed for the risk of outcome reporting bias

- Methotrexate monotherapy versus methotrexate combination therapy with non-biologic disease modifying antirheumatic drugs for rheumatoid arthritis (Review) [41]
- Antimalarials for treating rheumatoid arthritis (Review) [46]
- Azathioprine for treating rheumatoid arthritis (Review) [47]
- Auranofin versus placebo in rheumatoid arthritis (Review) [48]
- Cyclophosphamide for treating rheumatoid arthritis (Review) [49]
- Cyclosporine for treating rheumatoid arthritis (Review) [50]
- Injectable gold for rheumatoid arthritis (Review) [51]
- Methotrexate for treating rheumatoid arthritis (Review) [52]
- Penicillamine for treating rheumatoid arthritis (Review) [53]
- Sulfasalazine for treating rheumatoid arthritis (Review) [54]
- Leflunomide for the treatment of rheumatoid arthritis (Review) [55]
- Folic acid and folinic acid for reducing side effects in patients receiving methotrexate for rheumatoid arthritis (Review) [56]
- Adalimumab for treating rheumatoid arthritis (Review) [57]
- Abatacept for rheumatoid arthritis (Review) [58]
- Anakinra for rheumatoid arthritis (Review) [59]
- Certolizumab pegol (CDP870) for rheumatoid arthritis in adults (Review) [60]
- Etanercept for the treatment of rheumatoid arthritis (Review) [61]
- Infliximab for the treatment of rheumatoid arthritis (Review) [62]
- Golimumab for rheumatoid arthritis (Review) [63]
- Tocilizumab for rheumatoid arthritis (Review) [64]

- Effects of glucocorticoids on radiological progression in rheumatoid arthritis (Review) [65]

A.1 DMARDs for the treatment of Rheumatoid Arthritis

Table A.1 ORB matrix – Antimalarials for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Davis '91	Included	✓	✓	×(D)	×(H)	×(H)	×(F.R)	✓	✓
Clark '93		✓	✓	✓	×(C)	×(C)	×(FR)	✓	×(H)
Blackburn '95		✓	✓	✓	✓	✓	×(D)	✓	×(H)
HERA '95		✓	✓	✓	✓	✓	✓	✓	×(I)
Popert '61	Excluded	×(H)	×(H)	×(H)	×(H)	×(H)	×(C)	×(C)	×(C)
Scull '62		×(F)	×(F)	×(H)	×(H)	×(H)	×(H)	×(H)	×(F)

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.2 ORB matrix – Azathioprine for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Levy '72	Included	✓	×(FR)	×(H)	×(H)	×(H)	×(FR)	×(G)	×(G)
Urowitz '73		✓	✓	×(C)	×(E)	×(H)	×(FR)	✓	×(FR)
Woodland '81		✓	×(FR)	✓	×(FR)	×(B)	✓	✓	×(H)
Barnes '69	Excluded	×(H)	×(H)	×(H)	×(H)	×(H)	×(H)	×(E)	×(G)
Dixon '71		×(G)	×(G)	×(H)	×(H)	×(H)	×(A)	×(H)	×(H)
Pedersen '84		×(H)	×(H)	×(H)	×(H)	×(H)	×(H)	×(D)	×(H)

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.3 ORB matrix – Cyclophosphamide for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
CCC '70	Included	✓	✓	×(H)	×(C)	×(E)	×(FR)	✓	✓
Townes '76		✓	✓	×(D)	×(FR)	×(FR)	×(FR)	✓	×(D)

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.4 ORB matrix – Cyclosporine for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Dougados '88	Included	✓	✓	✓	×(FR)	×(FR)	✓	✓	×(H)
Tugwell '90		✓	✓	✓	✓	✓	✓	✓	×(H)
Forre '94		✓	✓	✓	✓	✓	✓	✓	✓

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.5 ORB matrix – Injectable Gold for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
ERC '60	Included	×(FR)	✓	×(FR)	✓	✓	×(FR)	✓	×(FR)
CCC '73		×(FR)	✓	×(H)	×(A)	×(E)	×(FR)	×(FR)	×(FR)
Sigler '74		×(C)	×(C)	×(H)	×(H)	×(H)	×(C)	×(C)	×(FR)
Ward '83		×(C)	✓	×(H)	✓	✓	×(C)	✓	×(H)

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.6 ORB matrix – Methotrexate for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Andersen '85	Included	✓	✓	✓	✓	✓	✓	✓	×(H)
Weinblatt '85		✓	✓	×(H)	✓	✓	✓	✓	×(H)
Williams '85		✓	✓	✓	✓	✓	✓	×(C)	×(H)
Furst '90		✓	✓	✓	✓	✓	✓	✓	×(H)
Pinheiro '93		✓	×(G)	✓	×(H)	×(H)	✓	✓	×(H)

TJC: Tender joint count; SJC: Swollen joint count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.7 ORB matrix – Penicillamine for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Andrews '73	Included	✓	×(C)	×(C)	✓	✓	✓	✓	×(A)
Dixon '75		×(G)	×(G)	✓	×(G)	×(G)	×(C)	✓	×(A)
Mery '76		✓	×(G)	×(G)	×(C)	×(C)	×(C)	✓	×(H)
Huskisson '76		×(C)	×(G)	×(C)	×(G)	×(G)	×(G)	×(C)	×(H)
Hamilton '77		×(G)	×(G)	×(G)	×(G)	×(G)	×(C)	×(C)	×(A)
Shiokawa '77		×(E)	×(E)	×(G)	×(G)	✓	×(E)	×(F)	×(C)
Williams '83		✓	✓	✓	✓	✓	×(FR)	✓	×(E)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.8 ORB matrix – Sulfasalazine for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Pullar '83	Included	×(E)	×(G)	×(E)	×(G)	×(G)	×(E)	×(E)	×(H)
Williams '88		✓	✓	×(G)	✓	✓	×(FR)	✓	×(H)
Skosey '88		×(A)	×(A)	×(C)	×(C)	×(C)	×(C)	×(C)	×(H)
Ebringer '92		✓	✓	✓	×(D)	×(D)	×(D)	✓	×(FR)
Danis '92		×(FR)	×(G)	×(FR)	×(G)	×(G)	×(FR)	×(FR)	×(H)
Hannonen '93		✓	✓	✓	✓	✓	×(C)	×(D)	✓
Farr '95		✓	×(G)	✓	×(D)	×(D)	×(FR)	✓	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.9 ORB matrix – Methotrexate monotherapy vs. Methotrexate combination therapy for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Willkins '92	Included	×(FR)	×(FR)	× (G)	×(FR)	×(FR)	×(D)	×(G)	×(H)
Ferraz '94		×(FR)	✓	✓	×(G)	×(G)	✓	✓	×(H)
Haagsma '94		×(FR)	×(FR)	✓	×(FR)	×(G)	×(FR)	✓	×(H)
Tugwell '95		✓	✓	✓	✓	✓	✓	✓	×(H)
Willkins '95		×(FR)	×(FR)	× (G)	×(FR)	×(FR)	×(D)	×(G)	×(H)
O'Dell '96		✓	✓	× (C)	✓	✓	×(C)	✓	×(G)
Haagsma '97		✓	×(FR)	✓	✓	×(D)	✓	✓	×(G)
Dougados '99		×(C)	×(C)	× (D)	×(C)	×(C)	×(C)	×(C)	×(C)
Hanyu '99		×(FR)	×(FR)	×(G)	×(G)	×(FR)	×(FR)	✓	×(G)
Islam '00		×(FR)	×(FR)	×(G)	×(FR)	×(FR)	×(FR)	×(FR)	×(H)
Kremer '02		✓	✓	✓	✓	✓	✓	✓	×(H)
Marchesoni '03		×(F)	×(F)	× (F)	×(F)	×(F)	×(F)	×(F)	✓
Ichikawa '05		×(F)	×(F)	× (F)	×(F)	×(F)	×(F)	×(F)	×(FR)
Lehman '05		✓	✓	✓	✓	✓	✓	✓	×(H)
Jarett '06		×(A)	×(A)	✓	✓	✓	×(G)	×(A)	✓
Hetland '06		×(E)	×(F)	× (F)	×(E)	×(E)	×(FR)	×(E)	×(FR)
O'Dell '06		×(C)	×(C)	× (C)	×(C)	×(C)	×(C)	×(C)	×(G)
Capell '07		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(A)
Ogrendik '07		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(H)
Calguneri '99	Excluded	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(G)	×(FR)	×(G)
Mottaghi '05		×(E)	×(E)	×(E)	×(E)	×(E)	×(E)	×(E)	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.10 ORB matrix – Leflunomide for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Rozman '94a	Included	✓	✓	×(G)	✓	✓	×(G)	×(D)	×(H)
Mladenovic '95		✓	✓	✓	✓	✓	✓	✓	×(H)
Emery '99		✓	✓	✓	✓	✓	✓	✓	✓
Smolen '99		✓	✓	✓	✓	✓	✓	✓	×(H)
Strand '99a		✓	✓	✓	✓	✓	✓	✓	✓
Bao '00		✓	✓	✓	✓	✓	×(E)	✓	×(H)
Sharp '00		×(E)	×(E)	×(E)	×(E)	×(E)	×(E)	×(E)	✓
Cohen '01		✓	✓	✓	✓	✓	✓	✓	✓
Hu '01		×(D)	×(A)	×(D)	×(D)	×(A)	×(D)	×(D)	×(H)
Kalden '01		×(E)	×(E)	×(E)	×(C)	×(C)	×(FR)	×(C)	×(D)
Larsen '01		×(E)	×(E)	×(E)	×(E)	×(E)	×(E)	×(C)	✓
Scott '01		✓	✓	✓	✓	✓	✓	✓	✓
Jakez-Ocampo'02		×(F)	×(F)	×(F)	×(F)	×(F)	×(F)	×(F)	×(G)
Kremer '02		✓	✓	✓	✓	✓	✓	✓	×(H)
Reece '02		✓	✓	✓	✓	✓	✓	✓	×(H)
Bao '03		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(H)
Kremer '04		✓	✓	✓	✓	✓	✓	✓	×(H)
Mariette '04		×(C)	×(C)	×(E)	×(E)	×(E)	×(FR)	×(C)	×(H)
Poor '04		✓	✓	×(E)	×(E)	×(E)	✓	×(E)	×(H)
Dougados '05		✓	✓	✓	×(A)	×(A)	✓	✓	×(G)
Antony '06		×(FR)	×(FR)	×(G)	×(FR)	×(FR)	×(FR)	×(FR)	×(H)
Karanikolas '06		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(G)
Fiehn '07		×(A)	×(A)	×(G)	×(E)	×(G)	×(G)	×(C)	×(H)
Wislowska '07		✓	✓	✓	×(E)	×(G)	✓	✓	×(H)
Grijalva '07	Excluded	×(H)	×(H)	×(H)	×(H)	×(H)	×(H)	×(H)	×(G)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.11 ORB matrix – Folic Acid for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Buckley '90	Included	×(C)	✓	×(C)	✓	×(C)	×(C)	×(C)	×(H)
Hanrahan '88		✓	✓	×(FR)	×(H)	×(H)	×(H)	×(H)	×(H)
Morgan '90		×(FR)	×(FR)	×(H)	×(FR)	×(FR)	×(D)	×(F)	×(H)
Joyce '91		✓	✓	×(C)	×(C)	×(H)	×(H)	×(C)	×(H)
Shiroky '93		✓	✓	×(H)	✓	×(FR)	×(FR)	×(FR)	×(H)
Weinblatt '93		✓	✓	×(H)	✓	×(FR)	×(H)	×(D)	×(H)
Morgan '94		✓	✓	×(H)	✓	×(FR)	×(FR)	×(A)	×(D)
Andersen '97	Excluded	×(G)	× (D)	× (D)	× (D)	× (D)	× (D)	× (D)	×(H)
Morgan '98		×(H)	×(H)	×(H)	×(H)	×(H)	×(H)	×(H)	×(G)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

A.2 Biologics for the treatment of Rheumatoid Arthritis

Table A.12 ORB matrix – Adalimumab for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Weinblatt '03	Included	✓	✓	✓	✓	✓	✓	✓	×(H)
Furst '03		×(F)	×(F)	×(F)	×(F)	×(F)	×(F)	×(F)	×(H)
Van de Putte '03		✓	✓	✓	✓	✓	✓	✓	×(H)
Van de Putte '04		✓	✓	✓	✓	✓	✓	×(C)	×(H)
Keystone '04		✓	✓	✓	✓	✓	✓	✓	✓
Rau '04		×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.13 ORB matrix – Anakinra for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Bresnihan '98	Included	×(F.R)	×(FR)	✓	×(FR)	×(FR)	✓	✓	✓
Cohen '02		×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(H)
Fleischman '03		×(H)	×(H)	×(H)	×(H)	×(H)	×(H)	× (H)	×(H)
Cohen '04		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	✓	✓	×(I)
Genovese '04		× (F)	× (F)	× (F)	× (F)	× (F)	× (F)	×(F)	×(I)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.14 ORB matrix – Abatacept for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Moreland '02	Included	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(H)
Kremer '03		✓	✓	✓	✓	✓	✓	×(C)	×(H)
Genovese '05		×(E)	×(E)	×(E)	×(E)	×(E)	✓	×(E)	×(H)
Kremer '06		✓	✓	✓	✓	✓	✓	×(FR)	✓
Weinblatt '06		×(E)	×(E)	✓	✓	✓	✓	×(E)	×(H)
Weinblatt '07		× (FR)	× (FR)	✓	×(FR)	×(FR)	✓	×(FR)	×(H)
Schiff '08		× (E)	× (E)	×(E)	×(E)	×(E)	✓	×(E)	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.15 ORB matrix – Golimumab for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Kay '08	Included	×(E)	×(E)	×(E)	×(E)	×(E)	×(E)	×(C)	×(H)
Kaystone '09		×(E)	×(E)	×(E)	×(E)	×(E)	✓	×(E)	×(H)
Emery '09		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(H)
Smolen '09		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	✓	×(FR)	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.16 ORB matrix – Certolizumab Pegol for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Choy '02	Included	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(H)
FAST4WARD '05		×(C)	×(C)	✓	×(C)	×(C)	✓	×(C)	×(H)
RAPID1 '05		×(C)	×(C)	✓	×(C)	×(C)	✓	×(C)	✓
RAPID2 '07		×(FR)	×(FR)	✓	×(FR)	×(FR)	✓	×(FR)	✓

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.17 ORB matrix – Tocilizumab for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Choy '02	Included	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(H)
Nishimoto '04		×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(C)	×(H)
Maini '06		×(B)	×(B)	×(E)	×(E)	×(E)	×(E)	×(C)	×(H)
Nishimoto '07		×(E)	×(E)	×(E)	×(E)	×(E)	×(C)	×(E)	×(FR)
Smolen '08		✓	✓	✓	✓	✓	✓	✓	×(H)
Genovese '08		✓	✓	×(E)	×(E)	×(E)	✓	✓	×(I)
Emery '08		✓	✓	×(D)	×(D)	×(D)	✓	✓	×(H)
Nishimoto '09		×(E)	×(E)	×(E)	×(E)	×(E)	×(C)	✓	×(H)
Straub '06	Excluded	×(E)	×(C)	×(H)	×(E)	×(H)	×(H)	×(C)	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

TableA.18 ORB matrix – Infliximab for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Maini '98	Included	✓	✓	×(D)	×(D)	×(D)	×(D)	✓	×(H)
Maini '99		✓	✓	✓	✓	✓	✓	✓	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

Table A.19 ORB matrix – Etanercept for treating rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Moreland '99	Included	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(H)
Weinblatt '99		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	✓	×(FR)	×(H)
Bathon '00 (ERA)		×(E)	×(E)	×(E)	×(E)	×(E)	✓	×(E)	✓
Klareskog '04 (TEMPO)		×(E)	×(E)	✓	×(E)	×(E)	✓	×(E)	✓
Combe '06		×(FR)	×(FR)	✓	×(FR)	×(FR)	✓	×(FR)	×(H)
Marcora '06		×(E)	×(E)	×(E)	×(E)	×(E)	✓	×(FR)	×(H)
Emery '08 (COMET)		×(E)	×(E)	×(FR)	×(E)	×(E)	✓	×(E)	✓
Hu '09		×(F)	×(F)	×(F)	×(F)	×(F)	×(F)	×(F)	×(H)
Kameda '10		×(FR)	×(FR)	×(G)	×(FR)	×(G)	✓	×(FR)	×(H)

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

A.3 Glucocorticoids for the treatment of Rheumatoid Arthritis

Table A.20 ORB matrix – Glucocorticoids on radiological progression in rheumatoid arthritis

Study	Eligibility	TJC	SJC	Pain	Pat. Global	Phy. Global	Function	APR	RD
Empire '55	Included	×(FR)	×(FR)	×(G)	×(C)	×(C)	×(G)	×(FR)	×(FR)
Empire '57		×(FR)	×(FR)	×(G)	×(C)	×(C)	×(G)	×(FR)	✓
Joint '59		×(C)	×(C)	×(G)	×(G)	×(G)	×(C)	×(C)	×(C)
Joint '60		×(C)	×(C)	×(G)	×(G)	×(G)	×(C)	×(C)	✓
Harris '83		×(FR)	×(FR)	×(FR)	×(C)	×(C)	×(C)	×(FR)	✓
Van Gestel '95		×(C)	×(C)	×(C)	×(C)	×(H)	×(C)	×(C)	✓
Kirwan '95		×(C)	×(D)	×(C)	×(G)	×(G)	×(C)	×(C)	✓
Schaardenburg '95		×(C)	×(C)	×(G)	×(G)	×(G)	×(C)	×(C)	✓
Boers '97		×(C)	×(G)	×(G)	×(G)	×(G)	×(C)	×(C)	✓
Hansen '99		×(D)	×(C)	×(G)	×(D)	×(D)	×(C)	×(C)	✓
Van Everdingen '02		×(FR)	×(FR)	×(FR)	×(G)	×(G)	×(F.R)	×(FR)	✓
Capell '04		×(FR)	×(G)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	✓
Choy '05		×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	×(FR)	✓
Svensson '05		×(F)	×(F)	×(F)	×(F)	×(H)	×(C)	×(C)	✓
Wassenberg '05		×(F)	×(F)	×(A)	×(A)	×(G)	×(A)	×(A)	✓
Goekoop '05		×(G)	×(G)	×(G)	×(G)	×(G)	×(FR)	×(D)	✓

TJC: Tender Joint Count; SJC: Swollen Joint Count; Pat.: Patient; Phy.: Physician; APR: acute phase reactant; RD: Radiological Damage; FR: Fully reported

APPENDIX B

SAS & STATA SCRIPTS FOR THE SIMULATION STUDY ORB SCENARIOS

B.1 Simulation of Individual Participant Data

RandNormal (N, Mean, Cov);

n = 5 studies

N

N1t = 30; N2t = 40; N3t = 50; N4t = 60; N5t = 70; N1p = 30; N2p = 40; N3p = 50; N4p = 60; N5p = 70;

Mean

Mean1t = {10.0, 8.0, 28.0}; Mean2t = {11.0, 9.0, 29.0}; Mean3t = {12.0, 10.0, 30.0};
Mean4t = {13.0, 11.0, 31.0}; Mean5t = {14.0, 12.0, 32.0}; Mean1p = {15.0, 11.0, 34.0};
Mean2p = {16.0, 12.0, 35.0}; Mean3p = {17.0, 13.0, 36.0}; Mean4p = {18.0, 14.0, 37.0};
Mean5p = {19.0, 15.0, 38.0};

Cov

- Patient-level correlation = 0

  Cov1t= {49 0 0, 0 25 0, 0 0 121};
  Cov2t = {64 0 0, 0 36 0, 0 0 144};
  Cov3t = {81 0 0, 0 49 0, 0 0 169};
  Cov4t = {100 0 0, 0 64 0, 0 0 196};
  Cov5t = {121 0 0, 0 81 0, 0 0 225};
  Cov1p = {64 0 0, 0 36 0, 0 0 144};
  Cov2p = {81 0 0, 0 49 0, 0 0 169};
  Cov3p = {100 0 0, 0 64 0, 0 0 196};
  Cov4p = {121 0 0, 0 81 0, 0 0 225};
  Cov5p = {144 0 0, 0 100 0, 0 0 256};

- Patient-level correlation = 0.2

  Cov1t = {49.0 7.0 15.4, 7.0 25.0 11.0, 15.4 11.0 121.0};
  Cov2t = {64.0 9.6 19.2, 9.6 36.0 14.4, 19.2 14.4 144.0};
  Cov3t = {81.0 12.6 23.4, 12.6 49.0 18.2, 23.4 18.2 169.0};
  Cov4t = {100.0 16.0 28.0, 16.0 64.0 22.4, 28.0 22.4 196.0};
  Cov5t = {121.0 19.8 33.0, 19.8 81.0 27.0, 33.0 27.0 225.0};
  Cov1p = {64.0 9.6 19.2, 9.6 36.0 14.4, 19.2 14.4 144.0};
  Cov2p = {81.0 12.6 23.4, 12.6 49.0 18.2, 23.4 18.2 169.0};
  Cov3p = {100.0 16.0 28.0, 16.0 64.0 22.4, 28.0 22.4 196.0};
  Cov4p = {121.0 19.8 33.0, 19.8 81.0 27.0, 33.0 27.0 225.0};
  Cov5p = {144.0 24.0 40.8, 24.0 100.0 34.0, 40.8 34.0 289.0};

- Patient-level correlation = 0.5

Cov1t = {49.0 17.5 38.5, 17.5 25.0 27.5, 38.5 27.5 121.0};
Cov2t = {64.0 24.0 48.0, 24.0 36.0 36.0, 48.0 36.0 144.0};
Cov3t = {81.0 31.5 58.5, 31.5 49.0 45.5, 58.5 45.5 169.0};
Cov4t = {100.0 40.0 70.0, 40.0 64.0 56.0, 70.0 56.0 196.0};
Cov5t = {121.0 49.5 82.5, 49.5 81.0 67.5, 82.5 67.5 225.0};
Cov1p = {64.0 24.0 48.0, 24.0 36.0 36.0, 48.0 36.0 144.0};
Cov2p = {81.0 31.5 58.5, 31.5 49.0 45.5, 58.5 45.5 169.0};
Cov3p = {100.0 40.0 70.0, 40.0 64.0 56.0, 70.0 56.0 196.0};
Cov4p = {121.0 49.5 82.5, 49.5 81.0 67.5, 82.5 67.5 225.0};
Cov5p = {144.0 60.0 102.0, 60.0 100.0 85.0, 102.0 85.0 289.0};

- Patient-level correlation = 0.8

Cov1t = {49.0 28.0 61.6, 28.0 25.0 44.0, 61.6 44.0 121.0};
Cov2t = {64.0 38.4 76.8, 38.4 36.0 57.6, 76.8 57.6 144.0};
Cov3t = {81.0 50.4 93.6, 50.4 49.0 72.8, 93.6 72.8 169.0};
Cov4t = {100.0 64.0 112.0, 64.0 64.0 89.6, 112.0 89.6 196.0};
Cov5t = {121.0 79.2 132.0, 79.2 81.0 108.0, 132.0 108.0 225.0};
Cov1p = {64.0 38.4 76.8, 38.4 36.0 57.6, 76.8 57.6 144.0};
Cov2p = {81.0 50.4 93.6, 50.4 49.0 72.8, 93.6 72.8 169.0};
Cov3p = {100.0 64.0 112.0, 64.0 64.0 89.6, 112.0 89.6 196.0};
Cov4p = {121.0 79.2 132.0, 79.2 81.0 108.0, 132.0 108.0 225.0};
Cov5p = {144.0 96.0 163.2, 96.0 100.0 136.0, 163.2 136.0 289.0};

- Patient-level correlation = RA_corr: 0.67 (outcome 1 & outcome 2), 0.55 (outcome 1 & outcome 3), 0.38

Cov1t = {49.0 23.3 42.3, 23.3 25.0 21.1, 42.3 21.1 121.0};
Cov2t = {64.0 32.0 52.7, 32.0 36.0 27.6, 52.7 27.6 144.0};
Cov3t = {81.0 42.0 64.2, 42.0 49.0 34.9, 64.2 34.9 169.0};
Cov4t = {100.0 53.3 76.9, 53.3 64.0 43.0, 76.9 43.0 196.0};
Cov5t = {121.0 66.0 90.6, 66.0 81.0 51.8, 90.6 51.8 225.0};
Cov1p = {64.0 32.0 52.7, 32.0 36.0 27.6, 52.7 27.6 144.0};
Cov2p = {81.0 42.0 64.2, 42.0 49.0 34.9, 64.2 34.9 169.0};
Cov3p = {100.0 53.3 76.9, 53.3 64.0 43.0, 76.9 43.0 196.0};
Cov4p = {121.0 66.0 90.6, 66.0 81.0 51.8, 90.6 51.8 225.0};
Cov5p = {144.0 80.0 112.0, 80.0 100.0 65.3, 112.0 65.3 289.0};

RandNormal (N, Mean, Cov);

n = 10 studies

N

N1t = 10; N2t = 20; N3t = 30; N4t = 40; N5t = 50; N6t = 60; N7t = 70; N8t = 80; N9t = 90; N10t = 100;
N1p = 10; N2p = 20; N3p = 30; N4p = 40; N5p = 50; N6p = 60; N7p = 70; N8p = 80; N9p = 90; N10p = 100;

Mean

Mean1t = {8.0, 8.0, 20.0}; Mean2t = {9.0, 9.0, 22.0}; Mean3t = {10.0, 9.75, 24.0};
Mean4t = {11.0, 10.0, 26.0}; Mean5t = {11.5, 11.0, 28.0}; Mean6t = {12.5, 12.0, 30.0};
Mean7t = {13.0, 13.0, 32.0}; Mean8t = {14.0, 13.25, 34.0}; Mean9t = {15.0, 14.0, 36.0};
Mean10t = {16.0, 15.0, 38.0};
Mean1p = {13.0, 11.0, 26.0}; Mean2p = {14.0, 12.0, 28.0}; Mean3p = {15.0, 12.75, 30.0};
Mean4p = {16.0, 13.0, 32.0}; Mean5p = {16.5, 14.0, 34.0}; Mean6p = {17.5, 15.0, 36.0};
Mean7p = {18.0, 16.0, 38.0}; Mean8p = {19.0, 16.25, 40.0}; Mean9p = {20.0, 17.0, 42.0};
Mean10p = {21.0, 18.0, 44.0};

COV

- Patient-level correlation = 0

  Cov1t= {49 0 0, 0 25 0, 0 0 121};
  Cov2t = {64 0 0, 0 36 0, 0 0 144};
  Cov3t = {81 0 0, 0 49 0, 0 0 169};
  Cov4t = {100 0 0, 0 64 0, 0 0 196};
  Cov5t = {121 0 0, 0 81 0, 0 0 225};
  Cov6t = {144 0 0, 0 100 0, 0 0 289};
  Cov7t = {169 0 0, 0 121 0, 0 0 324};
  Cov8t = {196 0 0, 0 144 0, 0 0 361};
  Cov9t = {225 0 0, 0 169 0, 0 0 400};
  Cov10t = {256 0 0, 0 196 0, 0 0 441};
  Cov1p = {64 0 0, 0 36 0, 0 0 144};
  Cov2p = {81 0 0, 0 49 0, 0 0 169};
  Cov3p = {100 0 0, 0 64 0, 0 0 196};
  Cov4p = {121 0 0, 0 81 0, 0 0 225};
  Cov5p = {144 0 0, 0 100 0, 0 0 256};
  Cov6p = {169 0 0, 0 121 0, 0 0 324};
  Cov7p = {196 0 0, 0 144 0, 0 0 361};
  Cov8p = {225 0 0, 0 169 0, 0 0 400};
  Cov9p = {256 0 0, 0 196 0, 0 0 441};
  Cov10p = {289 0 0, 0 225 0, 0 0 484};

- Patient-level correlation = 0.2

Cov1t = {49.0 7.0 15.4, 7.0 25.0 11.0, 15.4 11.0 121.0};
Cov2t = {64.0 9.6 19.2, 9.6 36.0 14.4, 19.2 14.4 144.0};
Cov3t = {81.0 12.6 23.4, 12.6 49.0 18.2, 23.4 18.2 169.0};
Cov4t = {100.0 16.0 28.0, 16.0 64.0 22.4, 28.0 22.4 196.0};
Cov5t = {121.0 19.8 33.0, 19.8 81.0 27.0, 33.0 27.0 225.0};
Cov6t = {144.0 24.0 40.8, 24.0 100.0 34.0, 40.8 34.0 289.0};
Cov7t = {169.0 28.6 46.8, 28.6 121.0 39.6, 46.8 39.6 324.0};
Cov8t = {196.0 33.6 53.2, 33.6 144.0 45.6, 53.2 45.6 361.0};
Cov9t = {225.0 39.0 60.0, 39.0 169.0 52.0, 60.0 52.0 400.0};
Cov10t = {256.0 44.8 67.2, 44.8 196.0 58.8, 67.2 58.8 441.0};
Cov1p = {64.0 9.6 19.2, 9.6 36.0 14.4, 19.2 14.4 144.0};
Cov2p = {81.0 12.6 23.4, 12.6 49.0 18.2, 23.4 18.2 169.0};
Cov3p = {100.0 16.0 28.0, 16.0 64.0 22.4, 28.0 22.4 196.0};
Cov4p = {121.0 19.8 33.0, 19.8 81.0 27.0, 33.0 27.0 225.0};
Cov5p = {144.0 24.0 40.8, 24.0 100.0 34.0, 40.8 34.0 289.0};
Cov6p = {169.0 28.6 46.8, 28.6 121.0 39.6, 46.8 39.6 324.0};
Cov7p = {196.0 33.6 53.2, 33.6 144.0 45.6, 53.2 45.6 361.0};
Cov8p = {225.0 39.0 60.0, 39.0 169.0 52.0, 60.0 52.0 400.0};
Cov9p = {256.0 44.8 67.2, 44.8 196.0 58.8, 67.2 58.8 441.0};
Cov10p = {289.0 51.0 74.8, 51.0 225.0 66.0, 74.8 66.0 484.0};

- Patient-level correlation = 0.5

Cov1t = {49.0 17.5 38.5, 17.5 25.0 27.5, 38.5 27.5 121.0};
Cov2t = {64.0 24.0 48.0, 24.0 36.0 36.0, 48.0 36.0 144.0};
Cov3t = {81.0 31.5 58.5, 31.5 49.0 45.5, 58.5 45.5 169.0};
Cov4t = {100.0 40.0 70.0, 40.0 64.0 56.0, 70.0 56.0 196.0};
Cov5t = {121.0 49.5 82.5, 49.5 81.0 67.5, 82.5 67.5 225.0};
Cov6t = {144.0 60.0 102.0, 60.0 100.0 85.0, 102.0 85.0 289.0};
Cov7t = {169.0 71.5 117.0, 71.5 121.0 99.0, 117.0 99.0 324.0};
Cov8t = {196.0 84.0 133.0, 84.0 144.0 114.0, 133.0 114.0 361.0};
Cov9t = {225.0 97.5 150.0, 97.5 169.0 130.0, 150.0 130.0 400.0};
Cov10t = {256.0 112.0 168.0, 112.0 196.0 147.0, 168.0 147.0 441.0};
Cov1p = {64.0 24.0 48.0, 24.0 36.0 36.0, 48.0 36.0 144.0};
Cov2p = {81.0 31.5 58.5, 31.5 49.0 45.5, 58.5 45.5 169.0};
Cov3p = {100.0 40.0 70.0, 40.0 64.0 56.0, 70.0 56.0 196.0};
Cov4p = {121.0 49.5 82.5, 49.5 81.0 67.5, 82.5 67.5 225.0};
Cov5p = {144.0 60.0 102.0, 60.0 100.0 85.0, 102.0 85.0 289.0};
Cov6p = {169.0 71.5 117.0, 71.5 121.0 99.0, 117.0 99.0 324.0};
Cov7p = {196.0 84.0 133.0, 84.0 144.0 114.0, 133.0 114.0 361.0};
Cov8p = {225.0 97.5 150.0, 97.5 169.0 130.0, 150.0 130.0 400.0};
Cov9p = {256.0 112.0 168.0, 112.0 196.0 147.0, 168.0 147.0 441.0};
Cov10p = {289.0 127.5 187.0, 127.5 225.0 165.0, 187.0 165.0 484.0};

- Patient-level correlation = 0.8

Cov1t = {49.0 28.0 61.6, 28.0 25.0 44.0, 61.6 44.0 121.0};
Cov2t = {64.0 38.4 76.8, 38.4 36.0 57.6, 76.8 57.6 144.0};
Cov3t = {81.0 50.4 93.6, 50.4 49.0 72.8, 93.6 72.8 169.0};
Cov4t = {100.0 64.0 112.0, 64.0 64.0 89.6, 112.0 89.6 196.0};
Cov5t = {121.0 79.2 132.0, 79.2 81.0 108.0, 132.0 108.0 225.0};
Cov6t = {144.0 96.0 163.2, 96.0 100.0 136.0, 163.2 136.0 289.0};
Cov7t = {169.0 114.4 187.2, 114.4 121.0 158.4, 187.2 158.4 324.0};
Cov8t = {196.0 134.4 212.8, 134.4 144.0 182.4, 212.8 182.4 361.0};
Cov9t = {225.0 156.0 240.0, 156.0 169.0 208.0, 240.0 208.0 400.0};
Cov10t = {256.0 179.2 268.8, 179.2 196.0 235.2, 268.8 235.2 441.0};
Cov1p = {64.0 38.4 76.8, 38.4 36.0 57.6, 76.8 57.6 144.0};
Cov2p = {81.0 50.4 93.6, 50.4 49.0 72.8, 93.6 72.8 169.0};
Cov3p = {100.0 64.0 112.0, 64.0 64.0 89.6, 112.0 89.6 196.0};
Cov4p = {121.0 79.2 132.0, 79.2 81.0 108.0, 132.0 108.0 225.0};
Cov5p = {144.0 96.0 163.2, 96.0 100.0 136.0, 163.2 136.0 289.0};
Cov6p = {169.0 114.4 187.2, 114.4 121.0 158.4, 187.2 158.4 324.0};
Cov7p = {196.0 134.4 212.8, 134.4 144.0 182.4, 212.8 182.4 361.0};
Cov8p = {225.0 156.0 240.0, 156.0 169.0 208.0, 240.0 208.0 400.0};
Cov9p = {256.0 179.2 268.8, 179.2 196.0 235.2, 268.8 235.2 441.0};
Cov10p = {289.0 204.0 299.2, 204.0 225.0 264.0, 299.2 264.0 484.0};

- Patient-level correlation = RA_corr: 0.67 (outcome 1 & outcome 2), 0.55 (outcome 1 & outcome 3), 0.38

Cov1t = {49.0 23.3 42.3, 23.3 25.0 21.1, 42.3 21.1 121.0};
Cov2t = {64.0 32.0 52.7, 32.0 36.0 27.6, 52.7 27.6 144.0};
Cov3t = {81.0 42.0 64.2, 42.0 49.0 34.9, 64.2 34.9 169.0};
Cov4t = {100.0 53.3 76.9, 53.3 64.0 43.0, 76.9 43.0 196.0};
Cov5t = {121.0 66.0 90.6, 66.0 81.0 51.8, 90.6 51.8 225.0};
Cov6t = {144.0 80.0 112.0, 80.0 100.0 65.3, 112.0 65.3 289.0};
Cov7t = {169.0 95.3 128.5, 95.3 121.0 76.0, 128.5 76.0 324.0};
Cov8t = {196.0 112.0 146.0, 112.0 144.0 87.5, 146.0 87.5 361.0};
Cov9t = {225.0 130.0 164.7, 130.0 169.0 99.8, 164.7 99.8 400.0};
Cov10t = {256.0 149.3 184.5, 149.3 196.0 112.9, 184.5 112.9 441.0};
Cov1p = {64.0 32.0 52.7, 32.0 36.0 27.6, 52.7 27.6 144.0};
Cov2p = {81.0 42.0 64.2, 42.0 49.0 34.9, 64.2 34.9 169.0};
Cov3p = {100.0 53.3 76.9, 53.3 64.0 43.0, 76.9 43.0 196.0};
Cov4p = {121.0 66.0 90.6, 66.0 81.0 51.8, 90.6 51.8 225.0};
Cov5p = {144.0 80.0 112.0, 80.0 100.0 65.3, 112.0 65.3 289.0};
Cov6p = {169.0 95.3 128.5, 95.3 121.0 76.0, 128.5 76.0 324.0};
Cov7p = {196.0 112.0 146.0, 112.0 144.0 87.5, 146.0 87.5 361.0};
Cov8p = {225.0 130.0 164.7, 130.0 169.0 99.8, 164.7 99.8 400.0};
Cov9p = {256.0 149.3 184.5, 149.3 196.0 112.9, 184.5 112.9 441.0};
Cov10p = {289.0 169.9 205.3, 169.9 225.0 126.7, 205.3 126.7 484.0};

B.2 First Stage IPD meta-analysis

SAS PROC MIXED

```
proc mixed cl method=reml data=work.Datasim2;
by Simulation Study;
class Outcome id Treated;
  ods listing close;
  ods output solutionf=Fixed;
model Y = Outcome1 Outcome2 Outcome3 Treated*Outcome1 Treated*Outcome2
Treated*Outcome3/ noint s cl covb corrb;
repeated Outcome/ type=un subject=id;
run;
```

B.3 Second Stage IPD meta-analysis

STATA

Univariate fixed-effects meta-analysis

```
mvmeta y S, var(y1) fixed
mvmeta y S, var(y2) fixed
mvmeta y S, var(y3) fixed
```

Multivariate fixed-effects meta-analysis

```
mvmeta y S, fixed
```

APPENDIX C

MULTIVARIATE META-ANALYSIS FIXED-EFFECTS SIMULATION RESULTS ORB SCENARIOS

In the following tables μ_1 refers to the outcome 1 (TJC), μ_2 refers to the outcome 2 (SJC) and finally μ_3 refers to the outcome 3 (pain).

Table C.1 Complete case

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength		
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
0	5	UFMA/MFMA	0.044	-0.0001	-0.024	0.852	0.666	1.213	94.0%	96.2%	94.4%	100.0%	99.1%	99.9%	0.729	0.444	1.473			
	10		0.015	-0.003	0.018	0.756	0.625	1.041	93.8%	93.2%	94.5%	100.0%	99.7%	99.9%	0.572	0.391	1.083			
0.2	5	UFMA**	0.044	0.005	-0.014	0.852	0.666	1.222	94.0%	96.2%	93.9%	100.0%	99.3%	99.6%	0.729	0.444	1.494	MFMA vs. UFMA		
		MFMA***	0.038	0.002	-0.017	0.845	0.660	1.212	93.2%	95.8%	93.6%	100.0%	99.4%	99.7%	0.715	0.435	1.469	1.8%	1.9%	1.7%
	10	UFMA	0.015	0.005	0.018	0.756	0.625	1.044	93.8%	93.2%	94.0%	100.0%	99.7%	100.0%	0.572	0.390	1.091	MFMA vs. UFMA		
		MFMA	0.019	-0.005	0.022	0.746	0.616	1.032	93.4%	92.1%	93.0%	100.0%	99.7%	100.0%	0.557	0.379	1.066	2.6%	2.8%	2.3%
0.5	5	UFMA	0.044	0.013	0.007	0.852	0.667	1.223	94.0%	96.0%	94.0%	100.0%	99.3%	99.8%	0.729	0.444	1.496	MFMA vs. UFMA		
		MFMA	0.037	0.011	0.001	0.844	0.660	1.211	93.3%	95.4%	93.3%	100.0%	99.5%	99.8%	0.714	0.435	1.467	2.0%	2.1%	1.9%
	10	UFMA	0.015	0.014	0.021	0.756	0.624	1.044	93.8%	94.6%	93.6%	100.0%	99.9%	100.0%	0.572	0.390	1.091	MFMA vs. UFMA		
		MFMA	0.018	0.0001	0.024	0.745	0.615	1.030	93.4%	94.0%	93.2%	100.0%	99.9%	100.0%	0.556	0.378	1.062	2.8%	3.1%	2.6%
0.8	5	UFMA	0.044	0.022	0.033	0.852	0.667	1.224	94.0%	95.3%	93.6%	100.0%	99.5%	99.8%	0.729	0.446	1.499	MFMA vs. UFMA		
		MFMA	0.035	0.020	0.023	0.840	0.657	1.206	93.1%	94.0%	93.0%	100.0%	99.5%	99.7%	0.707	0.432	1.456	2.8%	2.9%	2.9%
	10	UFMA	0.015	0.016	0.022	0.756	0.624	1.044	93.8%	94.8%	93.9%	100.0%	99.8%	100.0%	0.572	0.390	1.090	MFMA vs. UFMA		
		MFMA	0.015	0.005	0.023	0.741	0.611	1.023	92.9%	93.3%	93.1%	100.0%	99.8%	100.0%	0.550	0.373	1.048	3.9%	4.3%	3.9%
RA_corr*	5	UFMA	0.044	0.018	0.010	0.852	0.667	1.223	94.0%	95.8%	93.8%	100.0%	99.5%	99.8%	0.729	0.445	1.495	MFMA vs. UFMA		
		MFMA	0.036	0.016	0.005	0.844	0.660	1.211	93.3%	95.3%	93.8%	100.0%	99.6%	99.8%	0.713	0.436	1.468	2.0%	2.0%	1.8%
	10	UFMA	0.015	0.016	0.022	0.756	0.624	1.044	93.8%	94.4%	93.8%	100.0%	99.9%	100.0%	0.572	0.390	1.090	MFMA vs. UFMA 10		
		MFMA	0.017	0.003	0.029	0.745	0.615	1.031	93.1%	93.8%	93.0%	100.0%	99.9%	100.0%	0.556	0.378	1.063	2.9%	3.0%	2.5%

RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMA: Univariate fixed-effects meta-analysis ***MFMA: Multivariate fixed-effects meta-analysis

Table C.2 ORB: all studies are excluded that show benefit of alternative treatment irrespective of significance

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength		
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
0	5	UFMA/MFMA	0.008	-0.088	-0.136	0.856	0.675	1.225	94.7%	97.2%	95.4%	100.0%	99.9%	100.0%	0.732	0.464	1.520			
	10		-0.114	-0.273	-0.254	0.765	0.649	1.061	95.3%	94.1%	95.2%	100.0%	100.0%	100.0%	0.598	0.495	1.190			
0.2	5	UFMA**	0.008	-0.086	-0.139	0.856	0.676	1.236	94.7%	97.1%	94.9%	100.0%	99.8%	100.0%	0.732	0.464	1.547	MFMA vs. UFMA		
		MFMA***	0.002	-0.084	-0.134	0.848	0.669	1.225	93.9%	96.9%	94.8%	100.0%	99.8%	100.0%	0.719	0.455	1.518	1.8%	2.0%	1.8%
	10	UFMA	-0.114	-0.262	-0.258	0.765	0.648	1.065	95.3%	94.5%	95.0%	100.0%	99.9%	100.0%	0.598	0.489	1.201	MFMA vs. UFMA		
		MFMA	-0.111	-0.256	-0.248	0.755	0.638	1.052	95.0%	93.8%	94.6%	100.0%	99.9%	100.0%	0.583	0.473	1.169	2.5%	3.0%	2.4%
0.5	5	UFMA	0.008	-0.077	-0.113	0.856	0.676	1.236	94.7%	96.7%	95.3%	100.0%	99.7%	99.9%	0.732	0.463	1.540	MFMA vs. UFMA		
		MFMA	0.009	-0.054	-0.086	0.847	0.666	1.220	93.6%	96.4%	94.8%	100.0%	99.7%	99.9%	0.717	0.447	1.496	2.1%	2.8%	2.4%
	10	UFMA	-0.114	-0.260	-0.258	0.765	0.648	1.065	95.3%	94.1%	94.6%	100.0%	99.9%	100.0%	0.598	0.487	1.201	MFMA vs. UFMA		
		MFMA	-0.094	-0.204	-0.203	0.753	0.632	1.046	94.8%	94.1%	94.5%	100.0%	99.9%	100.0%	0.575	0.441	1.135	3.1%	4.8%	3.5%
0.8	5	UFMA	0.008	-0.065	-0.084	0.856	0.677	1.237	94.7%	96.2%	95.7%	100.0%	99.7%	99.9%	0.732	0.462	1.537	MFMA vs. UFMA		
		MFMA	0.011	-0.019	-0.037	0.843	0.661	1.212	93.9%	95.5%	94.7%	100.0%	99.6%	99.7%	0.710	0.437	1.470	3.0%	4.3%	3.8%
	10	UFMA	-0.114	-0.263	-0.275	0.765	0.648	1.066	95.3%	95.2%	95.4%	100.0%	99.9%	100.0%	0.598	0.489	1.212	MFMA vs. UFMA		
		MFMA	-0.085	-0.136	-0.162	0.748	0.622	1.035	94.3%	95.1%	94.3%	100.0%	99.9%	100.0%	0.567	0.405	1.098	4.3%	7.6%	5.6%
RA_corr*	5	UFMA	0.008	-0.066	-0.097	0.856	0.676	1.234	94.7%	96.7%	95.1%	100.0%	99.7%	99.9%	0.732	0.461	1.532	MFMA vs. UFMA		
		MFMA	0.012	-0.036	-0.077	0.846	0.666	1.220	93.9%	96.6%	94.9%	100.0%	99.7%	99.9%	0.716	0.444	1.493	2.2%	2.9%	2.2%
	10	UFMA	-0.114	-0.264	-0.269	0.765	0.648	1.066	95.3%	94.7%	95.0%	100.0%	99.9%	100.0%	0.598	0.490	1.208	MFMA vs. UFMA		
		MFMA	-0.092	-0.184	-0.210	0.752	0.630	1.048	94.8%	95.0%	94.4%	100.0%	99.9%	100.0%	0.574	0.431	1.142	3.3%	5.3%	3.3%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMA: Univariate fixed-effects meta-analysis ***MFMA: Multivariate fixed-effects meta-analysis

Table C.3 ORB: 15/20/30% data suppressed from TJC/SJC/Pain respectively if result non-significant

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength		
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
0	5	UFMA/MFMA	-0.092	-0.219	-0.700	0.882	0.727	1.403	93.3%	94.5%	93.1%	100.0%	99.2%	99.9%	0.786	0.577	2.457			
	10		-0.240	-0.001	-0.005	0.800	0.676	1.169	90.3%	93.0%	94.7%	100.0%	97.6%	98.8%	0.698	0.457	1.366			
0.2	5	UFMA**	-0.092	-0.208	-0.546	0.882	0.729	1.373	93.3%	95.0%	92.3%	100.0%	99.3%	99.7%	0.786	0.574	2.182	MFMA vs. UFMA		
		MFMA***	-0.099	-0.207	-0.531	0.874	0.720	1.356	93.1%	94.3%	92.5%	100.0%	99.4%	99.7%	0.774	0.561	2.122	1.7%	2.1%	2.1%
	10	UFMA	-0.240	-0.422	-0.816	0.800	0.740	1.226	90.3%	85.2%	85.1%	100.0%	99.7%	100.0%	0.698	0.725	2.168	MFMA vs. UFMA		
		MFMA	-0.235	-0.426	-0.776	0.791	0.728	1.206	90.3%	84.4%	84.1%	100.0%	99.7%	99.8%	0.681	0.712	2.057	2.3%	2.9%	2.8%
0.5	5	UFMA	-0.092	-0.204	-0.526	0.882	0.727	1.364	93.3%	94.7%	92.1%	100.0%	99.3%	99.9%	0.786	0.571	2.138	MFMA vs. UFMA		
		MFMA	-0.093	-0.172	-0.421	0.871	0.709	1.321	93.1%	94.3%	84.0%	100.0%	99.4%	99.8%	0.767	0.532	1.922	2.3%	4.0%	4.8%
	10	UFMA	-0.240	-0.399	-0.825	1.333	0.733	1.224	90.3%	86.4%	92.7%	100.0%	99.8%	99.9%	0.698	0.696	2.179	MFMA vs. UFMA		
		MFMA	-0.227	-0.361	-0.659	1.156	0.705	1.175	90.2%	86.4%	86.2%	100.0%	99.8%	99.7%	0.671	0.627	1.814	3.0%	5.5%	6.1%
0.8	5	UFMA	-0.092	-0.202	-0.490	0.882	0.730	1.368	93.3%	93.3%	93.0%	100.0%	99.5%	99.8%	0.786	0.573	2.112	MFMA vs. UFMA		
		MFMA	-0.093	-0.120	-0.270	0.867	0.692	1.279	92.7%	93.8%	93.7%	100.0%	99.4%	99.7%	0.761	0.494	1.710	3.0%	7.2%	9.1%
	10	UFMA	-0.240	-0.393	-0.855	0.800	0.733	1.230	90.3%	87.4%	84.4%	100.0%	99.7%	100.0%	0.698	0.691	2.243	MFMA vs. UFMA		
		MFMA	-0.225	-0.272	-0.499	0.783	0.673	1.125	90.3%	87.2%	87.3%	100.0%	99.5%	100.0%	0.663	0.526	1.516	4.1%	10.1%	11.5%
RA_corr*	5	UFMA	-0.092	-0.202	-0.508	0.882	0.728	1.369	93.3%	93.9%	92.1%	100.0%	99.5%	99.8%	0.786	0.570	2.131	MFMA vs. UFMA		
		MFMA	-0.091	-0.143	-0.401	0.869	0.703	1.327	93.0%	94.3%	92.7%	100.0%	99.4%	99.6%	0.764	0.515	1.923	2.5%	4.9%	4.6%
	10	UFMA	-0.240	-0.387	-0.816	0.800	0.728	1.218	90.3%	86.6%	84.6%	100.0%	99.7%	100.0%	0.698	0.680	2.149	MFMA vs. UFMA		
		MFMA	-0.225	-0.319	-0.644	0.786	0.692	1.172	89.8%	86.7%	86.6%	100.0%	99.6%	99.8%	0.668	0.580	1.788	3.3%	6.6%	5.8%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMA: Univariate fixed-effects meta-analysis ***MFMA: Multivariate fixed-effects meta-analysis

Table C.4 ORB: all studies are excluded that have non-significant results

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength		
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
0	5	UFMA/MFMA	-0.848	-1.172	-1.775	1.045	0.983	1.680	91.7%	86.0%	89.2%	100.0%	100.0%	100.0%	1.810	2.340	5.974			
	10		-1.694	-2.183	-2.764	1.064	1.189	1.647	68.2%	52.6%	61.7%	100.0%	99.8%	100.0%	4.002	6.180	10.351			
0.2	5	UFMA**	-0.848	-1.159	-1.786	1.045	0.984	1.696	91.7%	85.9%	88.2%	100.0%	100.0%	100.0%	1.810	2.311	6.066	MFMA vs. UFMA		
		MFMA***	-0.838	-1.139	-1.753	1.036	0.969	1.673	91.9%	85.9%	87.1%	100.0%	100.0%	98.7%	1.775	2.235	5.874	1.6%	3.0%	2.4%
	10	UFMA	-1.694	-2.169	-2.790	1.064	1.199	1.657	68.2%	52.4%	61.2%	100.0%	100.0%	100.0%	4.002	6.141	10.529	MFMA vs. UFMA		
		MFMA	-1.686	-2.136	-2.770	1.055	1.178	1.636	68.2%	52.2%	60.8%	100.0%	99.7%	100.0%	3.957	5.951	10.350	1.7%	3.3%	2.3%
0.5	5	UFMA	-0.848	-1.128	-1.795	1.045	0.987	1.689	91.7%	84.9%	88.2%	100.0%	99.9%	100.0%	1.810	2.246	6.077	MFMA vs. UFMA		
		MFMA	-0.802	-1.012	-1.612	1.018	0.923	1.602	91.7%	85.7%	88.7%	100.0%	99.4%	100.0%	1.679	1.877	5.167	4.4%	11.4%	8.8%
	10	UFMA	-1.694	-2.149	-2.770	1.064	1.185	1.644	68.2%	53.1%	60.0%	100.0%	100.0%	99.9%	4.002	6.024	10.377	MFMA vs. UFMA		
		MFMA	-1.645	-1.971	-2.635	1.031	1.088	1.556	68.1%	52.7%	60.9%	100.0%	99.7%	99.8%	3.767	5.070	9.366	5.8%	14.7%	9.6%
0.8	5	UFMA	-0.848	-1.129	-1.797	1.045	0.990	1.708	91.7%	85.7%	89.8%	100.0%	100.0%	100.0%	1.810	2.254	6.145	MFMA vs. UFMA		
		MFMA	-0.797	-0.838	-1.413	1.011	0.850	1.523	91.9%	88.8%	90.6%	100.0%	99.8%	99.9%	1.657	1.424	4.316	5.4%	23.1%	17.2%
	10	UFMA	-1.694	-2.142	-2.811	1.064	1.177	1.656	68.2%	54.4%	60.8%	100.0%	99.9%	100.0%	4.002	5.976	10.646	MFMA vs. UFMA		
		MFMA	-1.608	-1.679	-2.458	1.012	0.949	1.465	68.4%	56.5%	63.0%	100.0%	100.0%	100.0%	3.609	3.719	8.190	8.6%	31.8%	19.4%
RA_corr*	5	UFMA	-0.848	-1.137	-1.780	1.045	0.989	1.701	91.7%	85.2%	88.3%	100.0%	99.9%	100.0%	1.810	2.272	6.063	MFMA vs. UFMA		
		MFMA	-0.793	-0.950	-1.582	1.011	0.902	1.613	92.2%	85.7%	87.3%	100.0%	96.9%	98.1%	1.651	1.715	5.106	5.4%	15.2%	8.8%
	10	UFMA	-1.694	-2.115	-2.751	1.064	1.168	1.626	68.2%	54.2%	60.7%	100.0%	99.8%	100.0%	4.002	5.836	10.215	MFMA vs. UFMA		
		MFMA	-1.631	-1.847	-2.596	1.019	1.037	1.542	68.1%	55.4%	60.9%	100.0%	99.4%	99.7%	3.697	4.488	9.118	7.7%	19.5%	9.4%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMA: Univariate fixed-effects meta-analysis ***MFMA: Multivariate fixed-effects meta-analysis

Table C.5 ORB: all studies are excluded that have non-significant results and show benefit of alternative treatment

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength		
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
0	5	UFMA/MFMA	-0.848	-1.172	-1.775	1.045	0.983	1.680	91.7%	86.0%	89.2%	100.0%	100.0%	100.0%	1.810	2.340	5.974			
	10		-1.694	-2.185	-2.770	1.064	1.189	1.648	68.2%	52.6%	61.7%	100.0%	99.8%	100.0%	4.002	6.189	10.389			
0.2	5	UFMA**	-0.848	-1.160	-1.786	1.045	0.985	1.696	91.7%	85.9%	88.2%	100.0%	100.0%	100.0%	1.810	2.315	6.066	MFMA vs. UFMA		
		MFMA***	-0.838	-1.140	-1.754	1.036	0.969	1.673	91.9%	85.9%	88.2%	100.0%	100.0%	100.0%	1.775	2.239	5.875	1.6%	3.0%	2.4%
	10	UFMA	-1.694	-2.182	-2.800	1.064	1.200	1.657	68.2%	52.2%	61.1%	100.0%	100.0%	100.0%	4.002	6.200	10.586	MFMA vs. UFMA		
		MFMA	-1.688	-2.149	-2.781	1.055	1.179	1.637	68.1%	51.8%	60.7%	100.0%	99.7%	100.0%	3.962	6.008	10.414	1.7%	3.3%	2.3%
0.5	5	UFMA	-0.848	-1.130	-1.795	1.045	0.988	1.689	91.7%	85.0%	88.2%	100.0%	100.0%	100.0%	1.810	2.252	6.077	MFMA vs. UFMA		
		MFMA	-0.803	-1.015	-1.613	1.018	0.923	1.602	91.7%	85.7%	88.7%	100.0%	99.5%	100.0%	1.681	1.882	5.168	4.4%	11.4%	8.8%
	10	UFMA	-1.694	-2.163	-2.795	1.064	1.186	1.645	68.2%	51.3%	59.6%	100.0%	100.0%	100.0%	4.002	6.084	10.520	MFMA vs. UFMA		
		MFMA	-1.652	-1.986	-2.663	1.031	1.089	1.558	67.8%	52.1%	62.3%	100.0%	99.7%	99.9%	3.790	5.131	9.516	5.8%	14.7%	9.6%
0.8	5	UFMA	-0.848	-1.129	-1.797	1.045	0.990	1.708	91.7%	85.7%	89.8%	100.0%	100.0%	100.0%	1.810	2.254	6.145	MFMA vs. UFMA		
		MFMA	-0.797	-0.838	-1.413	1.011	0.850	1.523	91.9%	88.8%	90.6%	100.0%	99.8%	99.9%	1.657	1.424	4.316	5.4%	23.1%	17.2%
	10	UFMA	-1.694	-2.152	-2.816	1.064	1.178	1.657	68.2%	54.3%	60.8%	100.0%	100.0%	100.0%	4.002	6.019	10.677	MFMA vs. UFMA		
		MFMA	-1.614	-1.687	-2.469	1.012	0.950	1.466	68.2%	56.3%	62.9%	100.0%	100.0%	100.0%	3.631	3.748	8.247	8.6%	31.7%	19.3%
RA_corr*	5	UFMA	-0.848	-1.141	-1.780	1.045	0.990	1.701	91.7%	85.3%	88.3%	100.0%	100.0%	100.0%	1.810	2.281	6.063	MFMA vs. UFMA		
		MFMA	-0.794	-0.953	-1.583	1.011	0.902	1.613	92.2%	88.3%	88.6%	100.0%	99.8%	99.6%	1.654	1.722	5.109	5.4%	15.2%	8.8%
	10	UFMA	-1.694	-2.129	-2.770	1.064	1.169	1.627	68.2%	54.0%	60.6%	100.0%	100.0%	100.0%	4.002	5.898	10.320	MFMA vs. UFMA		
		MFMA	-1.640	-1.861	-2.616	1.019	1.038	1.543	67.7%	54.8%	60.8%	100.0%	99.4%	99.7%	3.730	4.539	9.226	7.7%	19.5%	9.4%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMA: Univariate fixed-effects meta-analysis ***MFMA: Multivariate fixed-effects meta-analysis

APPENDIX D

SAS & STATA SCRIPTS FOR THE SIMULATION STUDY MISSING DATA AT PATIENT LEVEL SCENARIOS

D.1 Simulation of IPD with Missing Data

MISSING COMPLETELY AT RANDOM

Number of studies equal to 5

- Scenario i: 20% MCAR in the IPD in outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if id<=6 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if id>=14 and id<22 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if id>=41 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if id<=12 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if id>=24 and id<38 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if id<=36 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if id>=54 and id<62 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if id>=91 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if id<=72 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if id>=94 and id<108 then x1=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 10

- Scenario i: 20% MCAR in the IPD for outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if id<=2 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if id>=9 and id<13 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if id>=25 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if id<=8 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if id>=17 and id<27 then x1=.;
end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250; if id>=49 then x1=.;
end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250; if id<=14 then x1=.;
end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250; if id>=20 and id<36 then x1=.;
end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250; if id>=73 then x1=.;
end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250; if id<=20 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if id<=12 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if id>=16 and id<20 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if id>=55 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if id<=48 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if id>=67 and id<77 then x1=.;
end; drop simulation; run;
data work.P6b; set work.P6; do simulation=1 to 250; if id>=109 then x1=.;
end; drop simulation; run;
data work.P7b; set work.P7; do simulation=1 to 250; if id<=84 then x1=.;
end; drop simulation; run;
data work.P8b; set work.P8; do simulation=1 to 250; if id>=100 and id<116 then x1=.;
end; drop simulation; run;
data work.P9b; set work.P9; do simulation=1 to 250; if id>=163 then x1=.;
end; drop simulation; run;
data work.P10b; set work.P10; do simulation=1 to 250; if id<=120 then x1=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 5

- Scenario ii: 40% MCAR in the IPD for outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if id<=12 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if id>=12 and id<28 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if id>=31 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if id<=24 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if id>=21 and id<49 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if id<=42 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if id>=52 and id<68 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if id>=81 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if id<=84 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if id>=91 and id<119 then x1=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 10

- Scenario ii: 40% MCAR in the IPD for outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if id<=4 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if id>=6 and id<14 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if id>=19 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if id>=25 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if id>=15 and id<35 then x1=.;
end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250; if id<=24 then x1=.;
end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250; if id>=21 and id<49 then x1=.;
end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250; if id<=32 then x1=.;
end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250; if id>=55 then x1=.;
end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250; if id<=40 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if id<=14 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if id>=26 and id<34 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if id>=49 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if id>=65 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if id>=65 and id<77 then x1=.;
end; drop simulation; run;
data work.P6b; set work.P6; do simulation=1 to 250; if id<=84 then x1=.;
end; drop simulation; run;
data work.P7b; set work.P7; do simulation=1 to 250; if id>=91 and id<119 then x1=.;
end; drop simulation; run;
data work.P8b; set work.P8; do simulation=1 to 250; if id<=112 then x1=.;
end; drop simulation; run;
data work.P9b; set work.P9; do simulation=1 to 250; if id>=145 then x1=.;
end; drop simulation; run;
data work.P10b; set work.P10; do simulation=1 to 250; if id<=140 then x1=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 5

- Scenario iii: 20% MCAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250;
if id<=6 then x1=.; if id>=11 and id<17 then x2=.; if id>=21 and id<27 then x3=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250;
if id<=8 then x1=.; if id>=14 and id<22 then x2=.; if id>=27 and id<35 then x3=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250;
if id<=10 then x1=.; if id>=17 and id<27 then x2=.; if id>=33 and id<43 then x3=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250;
if id<=12 then x1=.; if id>=20 and id<32 then x2=.; if id>=40 and id<52 then x3=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250;
if id<=14 then x1=.; if id>=24 and id<38 then x2=.; if id>=47 and id<61 then x3=.;
end; drop simulation; run;
 data work.P1b; set work.P1; do simulation=1 to 250;
if id<=36 then x1=.; if id>=41 and id<47 then x2=.; if id>=51 and id<57 then x3=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250;
if id<=48 then x1=.; if id>=54 and id<62 then x2=.; if id>=67 and id<75 then x3=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250;
if id<=60 then x1=.; if id>=67 and id<77 then x2=.; if id>=83 and id<93 then x3=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250;
if id<=72 then x1=.; if id>=80 and id<92 then x2=.; if id>=100 and id<112 then x3=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250;
if id<=84 then x1=.; if id>=94 and id<108 then x2=.; if id>=117 and id<131 then x3=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 10

- Scenario iii: 20% MCAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250;
if id<=2 then x1=.; if id=5 and id=6 then x2=.; if id>=9 then x3=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250;
if id<=4 then x1=.; if id>=9 and id<13 then x2=.; if id>=17 then x3=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250;
if id<=6 then x1=.; if id>=11 and id<17 then x2=.; if id>=21 and id<27 then x3=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250;
if id<=8 then x1=.; if id>=14 and id<22 then x2=.; if id>=27 and id<35 then x3=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250;
if id<=10 then x1=.; if id>=17 and id<27 then x2=.; if id>=33 and id<43 then x3=.;
end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250;
if id<=12 then x1=.; if id>=20 and id<32 then x2=.; if id>=40 and id<52 then x3=.;
end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250;
if id<=14 then x1=.; if id>=24 and id<38 then x2=.; if id>=47 and id<61 then x3=.;
end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250;
if id<=16 then x1=.; if id>=20 and id<36 then x2=.; if id>=40 and id<56 then x3=.;
end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250;
if id<=18 then x1=.; if id>=27 and id<45 then x2=.; if id>=45 and id<63 then x3=.;
end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250;
if id<=20 then x1=.; if id>=25 and id<45 then x2=.; if id>=45 and id<65 then x3=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 5

- Scenario iv: 40% MCAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250;
if id<=12 then x1=.; if id>=9 and id<21 then x2=.; if id>=19 then x3=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250;
if id<=16 then x1=.; if id>=12 and id<28 then x2=.; if id>=25 then x3=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250;
if id<=20 then x1=.; if id>=15 and id<35 then x2=.; if id>=31 then x3=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250;
if id<=24 then x1=.; if id>=18 and id<42 then x2=.; if id>=37 then x3=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250;
if id<=28 then x1=.; if id>=21 and id<49 then x2=.; if id>=43 then x3=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250;
if id<=42 then x1=.; if id>=39 and id<51 then x2=.; if id>=49 then x3=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250;
if id<=56 then x1=.; if id>=52 and id<68 then x2=.; if id>=65 then x3=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250;
if id<=70 then x1=.; if id>=65 and id<77 then x2=.; if id>=81 then x3=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250;
if id<=84 then x1=.; if id>=78 and id<102 then x2=.; if id>=97 then x3=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250;
if id<=98 then x1=.; if id>=91 and id<119 then x2=.; if id>=113 then x3=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 10

- Scenario iv: 40% MCAR in the IPD for all outcomes

```
data work.P1b; set work.P1; do simulation=1 to 250;
if id<=14 then x1=.; if id>=13 and id<17 then x2=.; if id>=17 then x3=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250;
if id<=28 then x1=.; if id>=26 and id<34 then x2=.; if id>=33 then x3=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250;
if id<=42 then x1=.; if id>=39 and id<51 then x2=.; if id>=49 then x3=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250;
if id<=56 then x1=.; if id>=52 and id<68 then x2=.; if id>=65 then x3=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250;
if id<=70 then x1=.; if id>=65 and id<77 then x2=.; if id>=81 then x3=.;
end; drop simulation; run;
data work.P6b; set work.P6; do simulation=1 to 250;
if id<=84 then x1=.; if id>=78 and id<102 then x2=.; if id>=97 then x3=.;
end; drop simulation; run;
data work.P7b; set work.P7; do simulation=1 to 250;
if id<=98 then x1=.; if id>=91 and id<119 then x2=.; if id>=113 then x3=.;
end; drop simulation; run;
data work.P8b; set work.P8; do simulation=1 to 250;
if id<=112 then x1=.; if id>=104 and id<136 then x2=.; if id>=129 then x3=.;
end; drop simulation; run;
data work.P9b; set work.P9; do simulation=1 to 250;
if id<=126 then x1=.; if id>=117 and id<153 then x2=.; if id>=145 then x3=.;
end; drop simulation; run;
data work.P10b; set work.P10; do simulation=1 to 250;
if id<=140 then x1=.; if id>=130 and id<170 then x2=.; if id>=161 then x3=.;
end; drop simulation; run;
```

MISSING COMPLETELY AT RANDOM

Number of studies equal to 10

- Scenario iv: 40% MCAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250;
if id<=4 then x1=.; if id>=3 and id<7 then x2=.; if id>=7 then x3=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250;
if id<=8 then x1=.; if id>=6 and id<14 then x2=.; if id>=13 then x3=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250;
if id<=12 then x1=.; if id>=9 and id<21 then x2=.; if id>=19 then x3=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250;
if id<=16 then x1=.; if id>=12 and id<28 then x2=.; if id>=25 then x3=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250;
if id<=20 then x1=.; if id>=15 and id<35 then x2=.; if id>=31 then x3=.;
end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250;
if id<=24 then x1=.; if id>=18 and id<42 then x2=.; if id>=37 then x3=.;
end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250;
if id<=28 then x1=.; if id>=21 and id<49 then x2=.; if id>=43 then x3=.;
end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250;
if id<=32 then x1=.; if id>=24 and id<56 then x2=.; if id>=49 then x3=.;
end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250;
if id<=36 then x1=.; if id>=27 and id<63 then x2=.; if id>=55 then x3=.;
end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250;
if id<=40 then x1=.; if id>=30 and id<70 then x2=.; if id>=61 then x3=.;
end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 5

- Scenario v: 20% MAR in the IPD for outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>16.0 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>17.7 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>19.5 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>21.4 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>23.0 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>21.8 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>23.7 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>25.4 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>27.2 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>29.1 then x1=.;
end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 10

- Scenario v: 20% MAR in the IPD for outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>14.2 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>15.7 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>17.5 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>19.5 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>20.6 then x1=.;
end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250; if x1>22.5 then x1=.;
end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250; if x1>23.8 then x1=.;
end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250; if x1>25.9 then x1=.;
end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250; if x1>27.6 then x1=.;
end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250; if x1>29.5 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>19.9 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>21.5 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>23.5 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>25.0 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>26.5 then x1=.;
end; drop simulation; run;
data work.P6b; set work.P6; do simulation=1 to 250; if x1>28.2 then x1=.;
end; drop simulation; run;
data work.P7b; set work.P7; do simulation=1 to 250; if x1>29.7 then x1=.;
end; drop simulation; run;
data work.P8b; set work.P8; do simulation=1 to 250; if x1>31.7 then x1=.;
end; drop simulation; run;
data work.P9b; set work.P9; do simulation=1 to 250; if x1>33.5 then x1=.;
end; drop simulation; run;
data work.P10b; set work.P10; do simulation=1 to 250; if x1>35.1 then x1=.;
end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 5

- Scenario vi : 40% MAR in the IPD for outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>11.8 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>13.0 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>14.2 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>15.6 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>16.7 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>17.1 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>18.5 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>19.6 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>20.8 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>22.0 then x1=.;
end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 10

- Scenario vi : 40% MAR in the IPD for outcome 1

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>10.0 then x1=.;
end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>10.9 then x1=.;
end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>12.3 then x1=.;
end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>13.5 then x1=.;
end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>14.3 then x1=.;
end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250; if x1>15.6 then x1=.;
end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250; if x1>16.2 then x1=.;
end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250; if x1>17.8 then x1=.;
end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250; if x1>18.8 then x1=.;
end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250; if x1>20.0 then x1=.;
end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>15.1 then x1=.;
end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>16.3 then x1=.;
end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>17.7 then x1=.;
end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>18.9 then x1=.;
end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>19.3 then x1=.;
end; drop simulation; run;
data work.P6b; set work.P6; do simulation=1 to 250; if x1>20.6 then x1=.;
end; drop simulation; run;
data work.P7b; set work.P7; do simulation=1 to 250; if x1>21.4 then x1=.;
end; drop simulation; run;
data work.P8b; set work.P8; do simulation=1 to 250; if x1>22.8 then x1=.;
end; drop simulation; run;
data work.P9b; set work.P9; do simulation=1 to 250; if x1>24.1 then x1=.;
end; drop simulation; run;
data work.P10b; set work.P10; do simulation=1 to 250; if x1>25.3 then x1=.;
end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 5

- Scenario vii : 20% MAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>16.0 then x1=.;
if x2>12.2 then x2=.; if x3>37.7 then x3=.; end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>17.7 then x1=.;
if x2>14.1 then x2=.; if x3>39.0 then x3=.; end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>19.5 then x1=.;
if x2>15.9 then x2=.; if x3>40.7 then x3=.; end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>21.4 then x1=.;
if x2>17.8 then x2=.; if x3>42.9 then x3=.; end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>23.0 then x1=.;
if x2>19.5 then x2=.; if x3>44.4 then x3=.; end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>21.8 then x1=.;
if x2>16.0 then x2=.; if x3>44.2 then x3=.; end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>23.7 then x1=.;
if x2>18.0 then x2=.; if x3>46.1 then x3=.; end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>25.4 then x1=.;
if x2>19.6 then x2=.; if x3>47.9 then x3=.; end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>27.2 then x1=.;
if x2>21.6 then x2=.; if x3>49.6 then x3=.; end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>29.1 then x1=.;
if x2>23.3 then x2=.; if x3>52.1 then x3=.; end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 10

- Scenario vii : 20% MAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>14.2 then x1=.;
if x2>12.3 then x2=.; if x3>29.8 then x3=.; end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>15.7 then x1=.;
if x2>14.1 then x2=.; if x3>32.5 then x3=.; end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>17.5 then x1=.;
if x2>15.7 then x2=.; if x3>34.9 then x3=.; end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>19.5 then x1=.;
if x2>16.8 then x2=.; if x3>37.5 then x3=.; end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>20.6 then x1=.;
if x2>18.7 then x2=.; if x3>40.6 then x3=.; end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250; if x1>22.5 then x1=.;
if x2>20.4 then x2=.; if x3>44.3 then x3=.; end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250; if x1>23.8 then x1=.;
if x2>22.1 then x2=.; if x3>47.0 then x3=.; end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250; if x1>25.9 then x1=.;
if x2>23.3 then x2=.; if x3>50.1 then x3=.; end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250; if x1>27.6 then x1=.;
if x2>24.9 then x2=.; if x3>52.9 then x3=.; end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250; if x1>29.5 then x1=.;
if x2>26.8 then x2=.; if x3>55.5 then x3=.; end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>19.9 then x1=.;
if x2>16.2 then x2=.; if x3>36.9 then x3=.; end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>21.5 then x1=.;
if x2>17.8 then x2=.; if x3>38.6 then x3=.; end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>23.5 then x1=.;
if x2>19.6 then x2=.; if x3>41.1 then x3=.; end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>25.0 then x1=.;
if x2>20.7 then x2=.; if x3>44.5 then x3=.; end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>26.5 then x1=.;
if x2>22.5 then x2=.; if x3>48.4 then x3=.; end; drop simulation; run;
data work.P6b; set work.P6; do simulation=1 to 250; if x1>28.2 then x1=.;
if x2>24.2 then x2=.; if x3>51.1 then x3=.; end; drop simulation; run;
data work.P7b; set work.P7; do simulation=1 to 250; if x1>29.7 then x1=.;
if x2>26.3 then x2=.; if x3>54.2 then x3=.; end; drop simulation; run;
data work.P8b; set work.P8; do simulation=1 to 250; if x1>31.7 then x1=.;
if x2>27.3 then x2=.; if x3>57.1 then x3=.; end; drop simulation; run;
data work.P9b; set work.P9; do simulation=1 to 250; if x1>33.5 then x1=.;
if x2>28.8 then x2=.; if x3>59.3 then x3=.; end; drop simulation; run;
data work.P10b; set work.P10; do simulation=1 to 250; if x1>35.1 then x1=.;
if x2>30.6 then x2=.; if x3>62.5 then x3=.; end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 5

- Scenario viii: 40% MAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>11.8 then x1=.;
if x2>9.3 then x2=.; if x3>30.8 then x3=.; end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>13.0 then x1=.;
if x2>10.6 then x2=.; if x3>32.0 then x3=.; end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>14.2 then x1=.;
if x2>11.9 then x2=.; if x3>33.0 then x3=.; end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>15.6 then x1=.;
if x2>13.1 then x2=.; if x3>34.6 then x3=.; end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>16.7 then x1=.;
if x2>14.1 then x2=.; if x3>35.6 then x3=.; end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>17.1 then x1=.;
if x2>12.4 then x2=.; if x3>37.1 then x3=.; end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>18.5 then x1=.;
if x2>13.8 then x2=.; if x3>38.5 then x3=.; end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>19.6 then x1=.;
if x2>15.0 then x2=.; if x3>39.6 then x3=.; end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>20.8 then x1=.;
if x2>16.2 then x2=.; if x3>40.7 then x3=.; end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>22.0 then x1=.;
if x2>17.6 then x2=.; if x3>42.2 then x3=.; end; drop simulation; run;
```

MISSING AT RANDOM

Number of studies equal to 10

- Scenario viii: 40% MAR in the IPD for all outcomes

```
data work.T1b; set work.T1; do simulation=1 to 250; if x1>10.0 then x1=.;
if x2>9.5 then x2=.; if x3>23.2 then x3=.; end; drop simulation; run;
data work.T2b; set work.T2; do simulation=1 to 250; if x1>10.9 then x1=.;
if x2>10.5 then x2=.; if x3>24.9 then x3=.; end; drop simulation; run;
data work.T3b; set work.T3; do simulation=1 to 250; if x1>12.3 then x1=.;
if x2>11.5 then x2=.; if x3>27.2 then x3=.; end; drop simulation; run;
data work.T4b; set work.T4; do simulation=1 to 250; if x1>13.5 then x1=.;
if x2>12.1 then x2=.; if x3>29.4 then x3=.; end; drop simulation; run;
data work.T5b; set work.T5; do simulation=1 to 250; if x1>14.3 then x1=.;
if x2>13.4 then x2=.; if x3>31.8 then x3=.; end; drop simulation; run;
data work.T6b; set work.T6; do simulation=1 to 250; if x1>15.6 then x1=.;
if x2>14.6 then x2=.; if x3>34.3 then x3=.; end; drop simulation; run;
data work.T7b; set work.T7; do simulation=1 to 250; if x1>16.2 then x1=.;
if x2>15.5 then x2=.; if x3>36.3 then x3=.; end; drop simulation; run;
data work.T8b; set work.T8; do simulation=1 to 250; if x1>17.8 then x1=.;
if x2>16.3 then x2=.; if x3>39.0 then x3=.; end; drop simulation; run;
data work.T9b; set work.T9; do simulation=1 to 250; if x1>18.8 then x1=.;
if x2>17.2 then x2=.; if x3>40.9 then x3=.; end; drop simulation; run;
data work.T10b; set work.T10; do simulation=1 to 250; if x1>20.0 then x1=.;
if x2>18.6 then x2=.; if x3>43.3 then x3=.; end; drop simulation; run;
data work.P1b; set work.P1; do simulation=1 to 250; if x1>15.1 then x1=.;
if x2>12.8 then x2=.; if x3>29.5 then x3=.; end; drop simulation; run;
data work.P2b; set work.P2; do simulation=1 to 250; if x1>16.3 then x1=.;
if x2>13.9 then x2=.; if x3>31.2 then x3=.; end; drop simulation; run;
data work.P3b; set work.P3; do simulation=1 to 250; if x1>17.7 then x1=.;
if x2>14.9 then x2=.; if x3>33.0 then x3=.; end; drop simulation; run;
data work.P4b; set work.P4; do simulation=1 to 250; if x1>18.9 then x1=.;
if x2>15.2 then x2=.; if x3>35.7 then x3=.; end; drop simulation; run;
data work.P5b; set work.P5; do simulation=1 to 250; if x1>19.3 then x1=.;
if x2>16.5 then x2=.; if x3>38.2 then x3=.; end; drop simulation; run;
data work.P6b; set work.P6; do simulation=1 to 250; if x1>20.6 then x1=.;
if x2>17.8 then x2=.; if x3>40.6 then x3=.; end; drop simulation; run;
data work.P7b; set work.P7; do simulation=1 to 250; if x1>21.4 then x1=.;
if x2>19.1 then x2=.; if x3>42.7 then x3=.; end; drop simulation; run;
data work.P8b; set work.P8; do simulation=1 to 250; if x1>22.8 then x1=.;
if x2>19.6 then x2=.; if x3>45.0 then x3=.; end; drop simulation; run;
data work.P9b; set work.P9; do simulation=1 to 250; if x1>24.1 then x1=.;
if x2>20.4 then x2=.; if x3>47.2 then x3=.; end; drop simulation; run;
data work.P10b; set work.P10; do simulation=1 to 250; if x1>25.3 then x1=.;
if x2>21.8 then x2=.; if x3>49.5 then x3=.; end; drop simulation; run;
```

D.2 First Stage IPD meta-analysis

SAS

Proc mixed separate analysis MODEL

```
proc mixed cl method=reml data=work.Datasim2;
by Simulation Study;
class Outcome id Treated;
ods listing close;
ods output solutionf=Fixed1;
model Y = Outcome1 Treated*Outcome1/ noint s cl covb corrb;
repeated Outcome/ type=un subject=id;
run;

proc mixed cl method=reml data=work.Datasim2;
by Simulation Study;
class Outcome id Treated;
  ods listing close;
  ods output solutionf=Fixed2;
model Y = Outcome2 Treated*Outcome2/ noint s cl covb corrb;
repeated Outcome/ type=un subject=id;
run;

proc mixed cl method=reml data=work.Datasim2;
by Simulation Study;
class Outcome id Treated;
  ods listing close;
  ods output solutionf=Fixed3;
model Y = Outcome3 Treated*Outcome3/ noint s cl covb corrb;
repeated Outcome/ type=un subject=id;
run;
```

Proc mixed joint analysis MODEL

```
proc mixed cl method=reml data=work.Datasim2;
by Simulation Study;
class Outcome id Treated;
  ods listing close;
  ods output solutionf=Fixed;
model Y = Outcome1 Outcome2 Outcome3 Treated*Outcome1 Treated*Outcome2 Treated*Outcome3/ noint s cl
covb corrb;
repeated Outcome/ type=un subject=id;
run;
```

D.3 Second Stage IPD meta-analysis

STATA

Univariate fixed-effects meta-analysis

```
mvmeta y S, var(y1) fixed
mvmeta y S, var(y2) fixed
mvmeta y S, var(y3) fixed
```

Multivariate fixed-effects meta-analysis

```
mvmeta y S, fixed
```

APPENDIX E

MULTIVARIATE META-ANALYSIS FIXED-EFFECTS SIMULATIONS RESULTS MISSING COMPLETLY AT RANDOM AND MISSING AT RANDOM SCENARIOS

In the following tables μ_1 refers to the outcome 1 (TJC), μ_2 refers to the outcome 2 (SJC) and finally μ_3 refers to the outcome 3 (pain).

Table E.1 Complete case

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength		
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3
0	5	UFMAj/MFMA	0.044	-0.0001	-0.024	0.852	0.666	1.213	94.0%	96.2%	94.4%	100.0%	99.1%	99.9%	0.729	0.444	1.473			
	10		0.015	-0.003	0.018	0.756	0.625	1.041	93.8%	93.2%	94.5%	100.0%	99.7%	99.9%	0.572	0.391	1.083			
0.2	5	UFMAj**	0.044	0.005	-0.014	0.852	0.666	1.222	94.0%	96.2%	93.9%	100.0%	99.3%	99.6%	0.729	0.444	1.494	MFMA vs. UFMA		
		MFMA***	0.038	0.002	-0.017	0.845	0.660	1.212	93.2%	95.8%	93.6%	100.0%	99.4%	99.7%	0.715	0.435	1.469	1.8%	1.9%	1.7%
	10	UFMAj	0.015	0.005	0.018	0.756	0.625	1.044	93.8%	93.2%	94.0%	100.0%	99.7%	100.0%	0.572	0.390	1.091	MFMA vs. UFMA		
		MFMA	0.019	-0.005	0.022	0.746	0.616	1.032	93.4%	92.1%	93.0%	100.0%	99.7%	100.0%	0.557	0.379	1.066	2.6%	2.8%	2.3%
0.5	5	UFMAj	0.044	0.013	0.007	0.852	0.667	1.223	94.0%	96.0%	94.0%	100.0%	99.3%	99.8%	0.729	0.444	1.496	MFMA vs. UFMA		
		MFMA	0.037	0.011	0.001	0.844	0.660	1.211	93.3%	95.4%	93.3%	100.0%	99.5%	99.8%	0.714	0.435	1.467	2.0%	2.1%	1.9%
	10	UFMAj	0.015	0.014	0.021	0.756	0.624	1.044	93.8%	94.6%	93.6%	100.0%	99.9%	100.0%	0.572	0.390	1.091	MFMA vs. UFMA		
		MFMA	0.018	0.0001	0.024	0.745	0.615	1.030	93.4%	94.0%	93.2%	100.0%	99.9%	100.0%	0.556	0.378	1.062	2.8%	3.1%	2.6%
0.8	5	UFMAj	0.044	0.022	0.033	0.852	0.667	1.224	94.0%	95.3%	93.6%	100.0%	99.5%	99.8%	0.729	0.446	1.499	MFMA vs. UFMA		
		MFMA	0.035	0.020	0.023	0.840	0.657	1.206	93.1%	94.0%	93.0%	100.0%	99.5%	99.7%	0.707	0.432	1.456	2.8%	2.9%	2.9%
	10	UFMAj	0.015	0.016	0.022	0.756	0.624	1.044	93.8%	94.8%	93.9%	100.0%	99.8%	100.0%	0.572	0.390	1.090	MFMA vs. UFMA		
		MFMA	0.015	0.005	0.023	0.741	0.611	1.023	92.9%	93.3%	93.1%	100.0%	99.8%	100.0%	0.550	0.373	1.048	3.9%	4.3%	3.9%
RA_corr*	5	UFMAj	0.044	0.018	0.010	0.852	0.667	1.223	94.0%	95.8%	93.8%	100.0%	99.5%	99.8%	0.729	0.445	1.495	MFMA vs. UFMA		
		MFMA	0.036	0.016	0.005	0.844	0.660	1.211	93.3%	95.3%	93.8%	100.0%	99.6%	99.8%	0.713	0.436	1.468	2.0%	2.0%	1.8%
	10	UFMAj	0.015	0.016	0.022	0.756	0.624	1.044	93.8%	94.4%	93.8%	100.0%	99.9%	100.0%	0.572	0.390	1.090	MFMA vs. UFMA		
		MFMA	0.017	0.003	0.029	0.745	0.615	1.031	93.1%	93.8%	93.0%	100.0%	99.9%	100.0%	0.556	0.378	1.063	2.9%	3.0%	2.5%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.2 20% MCAR in one outcome (μ_1)

Patient level correlation	No. Studies	Method	Bias			SEsim			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.075	0.010	-0.006	0.946	0.665	1.214	94.4%	93.8%	93.6%	99.9%	99.3%	99.7%	0.900	0.443	1.473				
		UFMAs/MFMA	0.013	0.013	0.062	0.950	0.667	1.214	95.1%	93.4%	93.2%	99.7%	99.2%	99.3%	0.903	0.445	1.478				
	10	UFMAj/MFMA	0.043	0.003	-0.006	0.836	0.625	1.040	92.8%	94.4%	94.6%	99.9%	99.8%	99.9%	0.701	0.391	1.081				
		UFMAs/MFMA	0.035	0.049	-0.065	0.840	0.625	1.040	93.2%	93.9%	94.2%	99.9%	99.4%	99.8%	0.706	0.393	1.087				
0.2	5	UFMAj*	0.079	0.022	0.013	0.940	0.665	1.223	94.2%	94.1%	93.4%	99.9%	99.4%	99.5%	0.889	0.443	1.496	Comparison	μ_1	μ_2	μ_3
		UFMAs**	0.013	0.014	0.069	0.950	0.667	1.223	95.1%	93.7%	93.2%	99.7%	99.4%	99.2%	0.903	0.446	1.501	MFMA vs UFMAj	2.0%	2.0%	1.9%
		MFMA***	0.078	0.024	0.012	0.930	0.659	1.211	93.4%	93.9%	93.4%	99.9%	99.5%	99.4%	0.872	0.435	1.468	MFMA vs UFMAs	3.7%	2.2%	1.6%
	10	UFMAj	0.040	0.009	0.004	0.831	0.625	1.044	93.2%	94.2%	95.2%	99.9%	99.8%	99.9%	0.692	0.391	1.090				
		UFMAs	0.035	0.052	-0.047	0.840	0.625	1.045	93.2%	93.8%	95.1%	99.9%	99.5%	99.8%	0.706	0.394	1.094	MFMA vs UFMAj	2.9%	3.0%	2.5%
		MFMA	0.045	0.014	0.001	0.819	0.616	1.031	92.9%	92.9%	94.1%	99.9%	99.9%	99.9%	0.673	0.379	1.062	MFMA vs UFMAs	4.6%	2.7%	2.5%
0.5	5	UFMAj	0.082	0.040	0.044	0.915	0.666	1.223	94.6%	94.0%	94.5%	99.9%	99.6%	99.5%	0.845	0.445	1.498	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.013	0.017	0.072	0.950	0.668	1.224	95.1%	94.2%	94.4%	99.7%	99.5%	99.5%	0.903	0.447	1.503	MFMA vs UFMAj	2.1%	2.2%	2.1%
		MFMA	0.082	0.042	0.050	0.906	0.658	1.210	94.2%	93.8%	93.9%	99.9%	99.5%	99.6%	0.827	0.435	1.467	MFMA vs UFMAs	8.8%	2.5%	1.9%
	10	UFMAj	0.036	0.019	0.017	0.810	0.625	1.044	93.6%	93.9%	95.0%	99.9%	99.7%	99.9%	0.658	0.391	1.091				
		UFMAs	0.035	0.051	-0.020	0.840	0.626	1.046	93.2%	94.1%	94.7%	99.9%	99.5%	99.9%	0.706	0.394	1.094	MFMA vs UFMAj	3.1%	3.2%	2.8%
		MFMA	0.041	0.023	0.022	0.798	0.615	1.030	93.1%	92.7%	94.5%	99.9%	99.9%	99.9%	0.638	0.379	1.061	MFMA vs UFMAs	9.5%	3.1%	2.8%

*UFMAj: Univariate fixed-effects meta-analysis joint model **UFMAs: Univariate fixed-effects meta-analysis separate model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.2 20% MCAR in one outcome (μ1)

Patient level correlation	No. Studies	Method	Bias			SEsim			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0.8	5	UFMAj**	0.083	0.056	0.079	0.879	0.666	1.222	95.1%	95.4%	95.5%	100.0%	99.9%	99.8%	0.780	0.447	1.500	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.012	0.021	0.067	0.950	0.668	1.225	95.2%	95.3%	95.5%	99.8%	99.7%	99.9%	0.903	0.447	1.504	MFMA vs UFMAj	2.9%	3.1%	3.0%
		MFMA****	0.084	0.058	0.090	0.867	0.656	1.204	94.3%	95.0%	95.0%	99.9%	99.5%	99.7%	0.758	0.433	1.458	MFMA vs UFMAs	16.5%	3.3%	3.0%
	10	UFMAj	0.031	0.028	0.032	0.780	0.625	1.044	93.9%	94.1%	94.5%	99.9%	99.6%	99.9%	0.609	0.391	1.092				
		UFMAs	0.035	0.043	0.008	0.840	0.626	1.046	93.2%	93.9%	94.5%	99.9%	99.9%	99.9%	0.706	0.394	1.095	MFMA vs UFMAj	4.1%	4.3%	3.9%
		MFMA	0.039	0.031	0.041	0.764	0.611	1.024	93.0%	92.9%	94.1%	99.9%	99.8%	99.9%	0.585	0.375	1.050	MFMA vs UFMAs	17.0%	4.4%	4.1%
RA_corr*	5	UFMAj	0.083	0.049	0.048	0.895	0.666	1.222	94.7%	94.2%	94.1%	99.9%	99.6%	99.5%	0.809	0.446	1.496	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.013	0.019	0.067	0.950	0.668	1.224	95.1%	94.0%	93.9%	99.7%	99.6%	99.4%	0.903	0.447	1.502	MFMA vs UFMAj	2.2%	2.2%	1.9%
		MFMA	0.083	0.050	0.049	0.886	0.659	1.210	94.4%	94.9%	93.7%	99.9%	99.6%	99.5%	0.791	0.437	1.467	MFMA vs UFMAs	12.8%	2.4%	1.8%
	10	UFMAj	0.028	0.048	-0.041	0.794	0.626	1.045	93.7%	93.9%	94.3%	99.9%	99.7%	99.9%	0.632	0.394	1.094				
		UFMAs	0.035	0.049	-0.041	0.840	0.626	1.045	93.2%	93.9%	94.3%	99.9%	99.7%	99.9%	0.706	0.394	1.094	MFMA vs UFMAj	3.1%	3.2%	2.6%
		MFMA	0.023	0.044	-0.034	0.782	0.616	1.032	92.5%	92.5%	93.0%	99.9%	99.6%	99.9%	0.612	0.381	1.065	MFMA vs UFMAs	13.3%	3.2%	2.6%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.3 40% MCAR in one outcome (μ_1)

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.027	0.013	0.062	1.083	0.667	1.214	94.7%	93.4%	93.2%	99.5%	99.2%	99.3%	1.173	0.445	1.478				
		UFMAs/MFMA	0.026	0.013	0.062	1.091	0.667	1.214	94.9%	93.4%	93.2%	99.5%	99.2%	99.3%	1.191	0.445	1.478				
	10	UFMAj/MFMA	-0.015	0.049	-0.065	0.954	0.625	1.040	92.2%	93.9%	94.2%	99.8%	99.4%	99.8%	0.910	0.393	1.087				
		UFMAs/MFMA	-0.012	0.049	-0.065	0.962	0.625	1.040	92.8%	93.9%	94.2%	99.8%	99.4%	99.8%	0.926	0.393	1.087				
0.2	5	UFMAj*	0.029	0.014	0.069	1.069	0.667	1.223	94.9%	93.7%	93.2%	99.6%	99.4%	99.2%	1.143	0.446	1.501	Comparison	μ_1	μ_2	μ_3
		UFMAs**	0.026	0.014	0.069	1.091	0.667	1.223	94.9%	93.7%	93.2%	99.5%	99.4%	99.2%	1.191	0.446	1.501	MFMA vs UFMAj	2.0%	2.0%	1.9%
		MFMA***	0.024	0.018	0.086	1.058	0.661	1.211	95.0%	93.1%	92.7%	99.4%	99.2%	99.3%	1.120	0.437	1.475	MFMA vs UFMAs	5.9%	2.0%	1.9%
	10	UFMAj	-0.015	0.052	-0.047	0.942	0.625	1.045	92.1%	93.8%	95.1%	99.8%	99.5%	99.8%	0.888	0.394	1.094				
		UFMAs	-0.012	0.052	-0.047	0.962	0.625	1.045	92.8%	93.8%	95.1%	99.8%	99.5%	99.8%	0.926	0.394	1.094	MFMA vs UFMAj	3.4%	3.2%	2.7%
		MFMA	-0.026	0.045	-0.043	0.926	0.615	1.031	91.2%	92.7%	94.1%	99.8%	99.6%	99.8%	0.858	0.380	1.064	MFMA vs UFMAs	7.4%	3.2%	2.7%
0.5	5	UFMAj	0.030	0.017	0.072	1.012	0.668	1.224	95.6%	94.2%	94.4%	99.7%	99.5%	99.5%	1.026	0.447	1.503	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.026	0.017	0.072	1.091	0.668	1.224	94.9%	99.5%	94.4%	94.2%	99.5%	99.5%	1.191	0.447	1.503	MFMA vs UFMAj	2.2%	2.2%	2.1%
		MFMA	0.024	0.022	0.087	1.001	0.661	1.211	94.5%	94.0%	93.5%	99.7%	99.5%	99.5%	1.002	0.437	1.474	MFMA vs UFMAs	15.7%	2.2%	2.1%
	10	UFMAj	-0.005	0.051	-0.020	0.894	0.626	1.046	92.4%	94.1%	94.7%	99.9%	99.5%	99.9%	0.799	0.394	1.094				
		UFMAs	-0.012	0.051	-0.020	0.962	0.626	1.046	92.8%	94.1%	94.7%	99.8%	99.5%	99.9%	0.926	0.394	1.094	MFMA vs UFMAj	3.6%	3.4%	2.9%
		MFMA	-0.014	0.043	-0.020	0.878	0.615	1.030	91.3%	92.6%	94.1%	99.9%	99.6%	99.9%	0.770	0.380	1.062	MFMA vs UFMAs	16.8%	3.4%	2.9%

*UFMAj: Univariate fixed-effects meta-analysis joint model **UFMAs: Univariate fixed-effects meta-analysis separate model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.3 40% MCAR in one outcome (μ1)

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0.8	5	UFMAj**	0.029	0.022	0.067	0.926	0.668	1.225	95.2%	95.3%	95.5%	99.9%	99.7%	99.9%	0.858	0.447	1.504	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.026	0.022	0.067	1.091	0.668	1.225	95.0%	95.6%	95.5%	95.3%	99.7%	99.9%	1.191	0.447	1.504	MFMA vs UFMAj	3.0%	3.1%	3.0%
		MFMA****	0.025	0.026	0.076	0.912	0.658	1.206	94.8%	94.0%	94.8%	99.8%	99.7%	99.8%	0.832	0.434	1.461	MFMA vs UFMAs	30.0%	3.1%	3.0%
	10	UFMAj	0.008	0.043	0.006	0.820	0.626	1.046	92.9%	93.8%	94.4%	99.9%	99.9%	99.9%	0.672	0.394	1.095				
		UFMAs	-0.012	0.043	0.008	0.962	0.626	1.046	92.8%	93.9%	94.5%	99.8%	99.9%	99.9%	0.926	0.394	1.095	MFMA vs UFMAj	4.5%	4.8%	4.1%
		MFMA	0.0004	0.032	-0.004	0.801	0.611	1.024	91.8%	93.1%	93.7%	99.9%	99.6%	99.9%	0.642	0.374	1.049	MFMA vs UFMAs	30.6%	4.7%	4.1%
RA_corr*	5	UFMAj	0.030	0.019	0.067	0.965	0.668	1.224	95.4%	94.0%	93.9%	99.8%	99.6%	99.4%	0.932	0.447	1.502	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.026	0.019	0.067	1.091	0.668	1.224	94.9%	94.0%	93.9%	99.5%	99.6%	99.4%	1.191	0.447	1.502	MFMA vs UFMAj	2.3%	2.2%	2.0%
		MFMA	0.025	0.024	0.082	0.954	0.661	1.211	94.8%	94.2%	94.0%	99.8%	99.6%	99.4%	0.910	0.438	1.474	MFMA vs UFMAs	23.5%	2.2%	2.0%
	10	UFMAj	0.004	0.048	-0.041	0.853	0.626	1.045	92.2%	93.9%	94.3%	99.9%	99.7%	99.9%	0.728	0.394	1.094				
		UFMAs	-0.012	0.048	-0.041	0.962	0.626	1.045	92.8%	93.9%	94.3%	99.8%	99.7%	99.9%	0.926	0.394	1.094	MFMA vs UFMAj	3.6%	3.4%	2.8%
		MFMA	-0.005	0.040	-0.040	0.838	0.615	1.031	91.7%	93.2%	93.5%	99.9%	99.5%	99.9%	0.702	0.380	1.064	MFMA vs UFMAs	24.2%	3.4%	2.8%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.4 20% MCAR in all outcomes

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.019	0.008	0.082	0.947	0.742	1.349	95.6%	92.4%	93.2%	99.8%	97.1%	99.1%	0.898	0.551	1.826				
		UFMAs/MFMA	0.018	0.009	0.089	0.951	0.745	1.353	95.5%	92.7%	93.9%	99.7%	97.0%	99.0%	0.904	0.555	1.840				
	10	UFMAj/MFMA	0.014	0.056	-0.083	0.836	0.686	1.154	93.8%	94.1%	94.6%	100.0%	98.3%	99.9%	0.699	0.474	1.338				
		UFMAs/MFMA	0.011	0.057	-0.087	0.838	0.688	1.158	94.2%	94.3%	94.9%	100.0%	98.5%	99.9%	0.703	0.477	1.348				
0.2	5	UFMAj*	0.015	0.011	0.088	0.941	0.738	1.350	95.3%	93.1%	93.1%	99.9%	97.2%	98.7%	0.887	0.545	1.831	Comparison	μ_1	μ_2	μ_3
		UFMAs**	0.018	0.011	0.094	0.951	0.745	1.363	95.5%	93.4%	93.5%	99.7%	97.0%	98.5%	0.904	0.555	1.868	MFMA vs UFMAj	1.9%	2.0%	1.9%
		MFMA***	0.012	0.012	0.098	0.932	0.730	1.338	94.4%	92.4%	92.9%	100.0%	97.4%	98.7%	0.870	0.533	1.799	MFMA vs UFMAs	3.8%	3.9%	3.8%
	10	UFMAj	0.020	0.058	-0.065	0.831	0.683	1.152	93.0%	94.0%	95.5%	100.0%	98.5%	99.9%	0.691	0.469	1.330				
		UFMAs	0.011	0.063	-0.068	0.838	0.688	1.162	94.1%	94.2%	95.4%	99.9%	98.3%	99.8%	0.703	0.478	1.356	MFMA vs UFMAj	2.8%	3.2%	2.6%
		MFMA	0.020	0.051	-0.063	0.819	0.672	1.136	92.0%	92.3%	94.9%	100.0%	98.6%	99.9%	0.671	0.454	1.295	MFMA vs UFMAs	4.5%	4.7%	4.5%
0.5	5	UFMAj	0.013	0.018	0.090	0.912	0.716	1.309	94.8%	93.9%	93.8%	100.0%	98.4%	99.5%	0.831	0.512	1.722	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.017	0.016	0.088	0.951	0.745	1.365	95.5%	94.0%	93.6%	99.7%	97.4%	99.2%	0.904	0.556	1.871	MFMA vs UFMAj	2.1%	2.2%	2.1%
		MFMA	0.028	0.055	-0.034	0.811	0.668	1.124	92.4%	93.6%	95.5%	100.0%	99.2%	100.0%	0.659	0.449	1.265	MFMA vs UFMAs	9.9%	9.8%	10.0%
	10	UFMAj	0.010	0.021	0.106	0.902	0.708	1.295	94.3%	93.8%	93.6%	100.0%	98.5%	99.5%	0.814	0.501	1.688				
		UFMAs	0.011	0.065	-0.037	0.838	0.689	1.163	94.1%	94.0%	94.3%	99.9%	98.3%	99.6%	0.703	0.479	1.355	MFMA vs UFMAj	3.0%	3.6%	2.8%
		MFMA	0.027	0.049	-0.032	0.799	0.656	1.108	91.5%	92.6%	94.2%	100.0%	99.0%	100.0%	0.639	0.432	1.229	MFMA vs UFMAs	9.2%	9.4%	9.2%

*UFMAj: Univariate fixed-effects meta-analysis joint model **UFMAs: Univariate fixed-effects meta-analysis separate model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.4 20% MCAR in all outcomes

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0.8	5	UFMAj**	0.017	0.024	0.080	0.879	0.689	1.263	95.1%	95.1%	95.1%	100.0%	99.2%	99.8%	0.773	0.476	1.603	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.018	0.021	0.067	0.951	0.745	1.366	95.6%	94.2%	94.9%	99.8%	97.4%	99.3%	0.904	0.555	1.872	MFMA vs UFMAj	2.7%	2.7%	2.8%
		MFMA****	0.016	0.025	0.093	0.867	0.680	1.245	94.5%	93.5%	94.5%	100.0%	99.1%	99.8%	0.751	0.463	1.559	MFMA vs UFMAs	16.8%	16.6%	16.9%
	10	UFMAj	0.033	0.045	-0.001	0.781	0.645	1.081	93.0%	94.5%	94.6%	100.0%	99.6%	100.0%	0.611	0.419	1.169				
		UFMAs	0.011	0.059	-0.003	0.838	0.689	1.164	94.2%	93.5%	94.1%	100.0%	98.6%	99.8%	0.703	0.479	1.354	MFMA vs UFMAj	3.8%	4.8%	3.6%
		MFMA	0.028	0.039	-0.003	0.766	0.630	1.062	91.7%	92.9%	94.2%	100.0%	99.6%	100.0%	0.588	0.398	1.127	MFMA vs UFMAs	16.5%	16.6%	16.7%
RA_corr*	5	UFMAj	0.015	0.022	0.088	0.894	0.708	1.312	94.9%	94.4%	94.5%	100.0%	98.6%	99.4%	0.800	0.502	1.730	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.018	0.018	0.081	0.951	0.745	1.364	95.6%	94.1%	94.2%	99.8%	97.5%	99.3%	0.904	0.556	1.868	MFMA vs UFMAj	2.2%	2.2%	2.0%
		MFMA	0.014	0.024	0.103	0.884	0.701	1.299	94.4%	94.2%	94.3%	100.0%	98.5%	99.4%	0.782	0.491	1.698	MFMA vs UFMAs	13.5%	11.6%	9.3%
	10	UFMAj	0.032	0.054	-0.056	0.795	0.662	1.127	93.0%	93.9%	95.7%	100.0%	99.1%	99.9%	0.633	0.441	1.273				
		UFMAs	0.011	0.062	-0.056	0.838	0.689	1.163	94.1%	93.3%	94.8%	99.9%	98.2%	99.7%	0.703	0.479	1.356	MFMA vs UFMAj	3.0%	3.6%	2.7%
		MFMA	0.030	0.051	-0.051	0.783	0.650	1.111	91.8%	92.8%	95.2%	100.0%	98.7%	99.9%	0.614	0.425	1.238	MFMA vs UFMAs	12.6%	10.9%	8.5%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.5 40% MCAR in all outcomes

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.032	0.019	0.078	1.078	0.840	1.537	94.4%	92.1%	93.6%	99.2%	93.2%	96.8%	1.164	0.706	2.368				
		UFMAs/MFMA	0.028	0.019	0.078	1.092	0.849	1.557	95.1%	92.7%	93.7%	99.2%	93.2%	96.7%	1.193	0.720	2.430				
	10	UFMAj/MFMA	-0.008	0.066	-0.074	0.951	0.781	1.305	91.6%	92.1%	93.3%	99.9%	94.4%	99.2%	0.905	0.614	1.708				
		UFMAs/MFMA	-0.009	0.067	-0.076	0.968	0.795	1.331	93.7%	93.4%	94.4%	99.9%	94.4%	99.4%	0.938	0.637	1.778				
0.2	5	UFMAj*	0.033	0.021	0.074	1.069	0.833	1.534	94.4%	91.5%	93.6%	99.4%	93.5%	97.0%	1.143	0.694	2.357	Comparison	μ_1	μ_2	μ_3
		UFMAs**	0.028	0.020	0.077	1.092	0.849	1.568	95.1%	92.6%	93.8%	99.2%	93.6%	96.8%	1.193	0.721	2.464	MFMA vs UFMAj	1.9%	2.2%	1.8%
		MFMA***	0.035	0.028	0.083	1.058	0.824	1.519	93.9%	91.2%	92.9%	99.3%	93.4%	97.0%	1.121	0.679	2.316	MFMA vs UFMAs	6.0%	5.8%	6.1%
	10	UFMAj	-0.026	0.038	-0.041	0.942	0.773	1.299	91.4%	92.7%	92.9%	99.9%	96.2%	99.2%	0.888	0.600	1.690				
		UFMAs	-0.007	0.067	-0.043	0.968	0.796	1.337	93.7%	93.9%	94.3%	99.9%	94.8%	99.4%	0.938	0.637	1.789	MFMA vs UFMAj	7.3%	7.9%	8.4%
		MFMA	-0.002	0.007	-0.026	0.900	0.736	1.233	83.8%	84.7%	79.9%	98.1%	94.0%	95.9%	0.810	0.541	1.522	MFMA vs UFMAs	12.3%	13.1%	13.6%
0.5	5	UFMAj	0.032	0.026	0.067	1.023	0.797	1.469	94.7%	92.6%	93.4%	99.6%	94.7%	98.1%	1.048	0.636	2.163	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.028	0.023	0.071	1.092	0.849	1.569	95.1%	93.5%	94.0%	99.2%	92.5%	97.0%	1.193	0.721	2.467	MFMA vs UFMAj	2.2%	2.4%	2.1%
		MFMA	0.031	0.034	0.076	1.012	0.788	1.453	93.3%	91.0%	92.9%	99.5%	94.4%	98.1%	1.024	0.621	2.118	MFMA vs UFMAs	14.1%	13.9%	14.1%
	10	UFMAj	0.014	0.037	-0.065	0.901	0.741	1.246	92.4%	93.3%	92.6%	99.8%	97.2%	99.7%	0.813	0.550	1.556				
		UFMAs	-0.007	0.058	0.004	0.968	0.796	1.338	93.7%	94.5%	93.9%	99.9%	95.5%	99.2%	0.938	0.637	1.790	MFMA vs UFMAj	7.7%	8.3%	8.1%
		MFMA	-0.005	0.024	-0.087	0.862	0.706	1.189	83.1%	84.0%	81.0%	98.6%	94.1%	97.8%	0.743	0.499	1.422	MFMA vs UFMAs	19.8%	20.4%	20.2%

*UFMAj: Univariate fixed-effects meta-analysis joint model **UFMAs: Univariate fixed-effects meta-analysis separate model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.5 40% MCAR in all outcomes

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0.8	5	UFMAj**	0.027	0.028	0.065	0.938	0.732	1.349	94.5%	93.9%	94.8%	99.8%	97.7%	99.0%	0.881	0.537	1.823	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.028	0.030	0.061	1.092	0.848	1.570	95.2%	94.5%	95.2%	99.3%	92.9%	96.7%	1.193	0.721	2.469	MFMA vs UFMAj	2.8%	3.0%	2.7%
		MFMA****	0.024	0.033	0.072	0.925	0.721	1.330	93.6%	92.6%	93.5%	99.8%	97.8%	98.9%	0.856	0.521	1.775	MFMA vs UFMAs	28.1%	27.6%	28.1%
	10	UFMAj	0.015	0.055	-0.001	0.830	0.682	1.146	93.2%	93.3%	93.9%	100.0%	98.7%	100.0%	0.689	0.469	1.313				
		UFMAs	-0.007	0.042	0.047	0.968	0.797	1.339	93.8%	94.2%	94.4%	100.0%	95.0%	99.3%	0.938	0.637	1.794	MFMA vs UFMAj	9.4%	11.2%	8.6%
		MFMA	0.053	0.060	0.072	0.782	0.636	1.085	82.7%	80.9%	83.8%	98.2%	94.8%	97.3%	0.615	0.408	1.183	MFMA vs UFMAs	33.4%	34.9%	33.0%
RA_corr*	5	UFMAj	0.035	0.031	0.077	0.991	0.785	1.478	94.8%	92.9%	93.3%	99.6%	95.2%	97.9%	0.983	0.618	2.191	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.028	0.027	0.078	1.092	0.849	1.569	95.1%	93.2%	94.3%	99.2%	92.5%	97.0%	1.193	0.721	2.467	MFMA vs UFMAj	2.4%	2.5%	2.0%
		MFMA	0.033	0.039	0.082	0.979	0.776	1.464	92.8%	92.3%	93.5%	99.7%	95.4%	98.0%	0.959	0.603	2.149	MFMA vs UFMAs	19.6%	16.4%	12.9%
	10	UFMAj	0.009	0.042	-0.011	0.875	0.731	1.253	92.7%	91.5%	92.1%	99.9%	97.2%	99.3%	0.765	0.536	1.571				
		UFMAs	-0.009	0.051	-0.018	0.968	0.797	1.338	93.6%	94.4%	93.9%	99.9%	95.3%	99.2%	0.938	0.637	1.790	MFMA vs UFMAj	8.4%	8.8%	8.4%
		MFMA	0.009	-0.004	0.020	0.831	0.692	1.190	81.8%	83.4%	81.2%	98.0%	94.6%	96.9%	0.691	0.478	1.416	MFMA vs UFMAs	25.2%	23.2%	19.6%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.6 20% MAR in one outcome (μ1)

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.301	0.013	0.062	0.724	0.667	1.214	94.0%	93.5%	93.3%	100.0%	99.3%	99.4%	0.615	0.445	1.478				
		UFMAs/MFMA	0.298	0.013	0.062	0.726	0.667	1.214	93.9%	93.5%	93.3%	100.0%	99.3%	99.4%	0.616	0.445	1.478				
	10	UFMAj/MFMA	0.356	0.049	-0.065	0.643	0.625	1.040	90.2%	94.0%	94.3%	100.0%	99.5%	99.9%	0.540	0.393	1.086				
		UFMAs/MFMA	0.356	0.049	-0.065	0.645	0.625	1.040	89.9%	94.0%	94.3%	100.0%	99.5%	99.9%	0.543	0.393	1.087				
0.2	5	UFMAj*	0.287	0.014	0.069	0.722	0.667	1.223	93.5%	93.8%	93.3%	100.0%	99.5%	99.3%	0.603	0.446	1.501	Comparison	μ_1	μ_2	μ_3
		UFMAs**	0.298	0.014	0.069	0.726	0.667	1.223	93.9%	93.8%	93.3%	100.0%	99.5%	99.3%	0.616	0.446	1.501	MFMA vs UFMAj	2.0%	2.0%	1.9%
		MFMA***	0.282	0.016	0.078	0.715	0.661	1.211	92.9%	93.2%	92.8%	100.0%	99.5%	99.3%	0.590	0.437	1.473	MFMA vs UFMAs	3.2%	2.0%	1.9%
	10	UFMAj	0.344	0.052	-0.047	0.641	0.625	1.045	90.1%	93.9%	95.2%	100.0%	99.6%	99.9%	0.529	0.394	1.094				
		UFMAs	0.356	0.052	-0.047	0.645	0.625	1.045	89.9%	93.9%	95.2%	100.0%	99.6%	99.9%	0.543	0.394	1.094	MFMA vs UFMAj	2.8%	3.0%	2.5%
		MFMA	0.341	0.049	-0.048	0.632	0.616	1.032	89.9%	93.0%	94.3%	100.0%	99.6%	100.0%	0.515	0.382	1.066	MFMA vs UFMAs	4.0%	3.0%	2.5%
0.5	5	UFMAj	0.228	0.017	0.072	0.719	0.668	1.224	94.7%	94.3%	94.5%	100.0%	99.6%	99.6%	0.569	0.447	1.503	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.298	0.017	0.072	0.726	0.668	1.224	93.9%	94.3%	94.5%	100.0%	99.6%	99.6%	0.616	0.447	1.503	MFMA vs UFMAj	2.1%	2.3%	2.2%
		MFMA	0.224	0.019	0.077	0.712	0.660	1.210	94.0%	93.4%	93.3%	100.0%	99.7%	99.5%	0.557	0.436	1.471	MFMA vs UFMAs	4.0%	2.3%	2.2%
	10	UFMAj	0.281	0.051	-0.020	0.639	0.626	1.046	91.6%	94.2%	94.8%	100.0%	99.6%	100.0%	0.487	0.394	1.094				
		UFMAs	0.356	0.051	-0.020	0.645	0.626	1.046	89.9%	94.2%	94.8%	100.0%	99.6%	100.0%	0.543	0.394	1.094	MFMA vs UFMAj	3.0%	3.4%	2.9%
		MFMA	0.277	0.044	-0.026	0.629	0.615	1.030	91.0%	92.6%	94.0%	100.0%	99.7%	100.0%	0.472	0.380	1.062	MFMA vs UFMAs	4.8%	3.4%	2.9%

*UFMAj: Univariate fixed-effects meta-analysis joint model **UFMAs: Univariate fixed-effects meta-analysis separate model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.6 20% MAR in one outcome (μ1)

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0.8	5	UFMAj**	0.119	0.022	0.067	0.754	0.668	1.225	96.0%	95.3%	95.5%	100.0%	99.7%	99.9%	0.583	0.447	1.504	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.298	0.022	0.067	0.744	0.668	1.225	93.9%	95.3%	95.5%	100.0%	99.7%	99.9%	0.616	0.447	1.504	MFMA vs UFMAj	2.7%	3.4%	3.3%
		MFMA****	0.120	0.025	0.074	0.726	0.657	1.204	95.4%	94.3%	94.9%	100.0%	99.7%	99.9%	0.568	0.432	1.456	MFMA vs UFMAs	4.9%	3.4%	3.3%
	10	UFMAj	0.281	0.051	-0.020	0.639	0.626	1.046	91.6%	94.2%	94.8%	100.0%	99.6%	100.0%	0.487	0.394	1.094				
		UFMAs	0.356	0.051	-0.020	0.645	0.626	1.046	89.9%	94.2%	94.8%	100.0%	99.6%	100.0%	0.543	0.394	1.094	MFMA vs UFMAj	3.0%	3.4%	2.9%
		MFMA	0.277	0.044	-0.026	0.629	0.615	1.030	91.0%	92.6%	94.0%	100.0%	99.7%	100.0%	0.472	0.380	1.062	MFMA vs UFMAs	4.8%	3.4%	2.9%
RA_corr*	5	UFMAj	0.162	0.019	0.067	0.731	0.668	1.224	95.9%	94.1%	94.0%	100.0%	99.7%	99.5%	0.560	0.447	1.502	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.298	0.019	0.067	0.726	0.668	1.224	93.9%	94.1%	94.0%	100.0%	99.7%	99.5%	0.616	0.447	1.502	MFMA vs UFMAj	2.1%	2.4%	2.1%
		MFMA	0.160	0.022	0.075	0.723	0.660	1.211	95.4%	94.0%	94.1%	100.0%	99.7%	99.4%	0.548	0.436	1.471	MFMA vs UFMAs	0.9%	2.4%	2.1%
	10	UFMAj	0.209	0.048	-0.041	0.649	0.626	1.045	92.1%	94.0%	94.4%	100.0%	99.8%	100.0%	0.464	0.394	1.094				
		UFMAs	0.356	0.048	-0.041	0.645	0.626	1.045	89.9%	94.0%	94.4%	100.0%	99.8%	100.0%	0.543	0.394	1.094	MFMA vs UFMAj	2.9%	3.5%	2.8%
		MFMA	0.206	0.042	-0.045	0.639	0.615	1.031	91.9%	92.9%	93.4%	100.0%	99.8%	100.0%	0.451	0.380	1.064	MFMA vs UFMAs	1.8%	3.5%	2.8%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.7 40% MAR in one outcome (μ_1)

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.592	0.013	0.062	0.704	0.667	1.214	84.3%	93.5%	93.3%	100.0%	99.3%	99.4%	0.846	0.445	1.478				
		UFMAs/MFMA	0.592	0.013	0.062	0.709	0.667	1.214	84.9%	93.5%	93.3%	100.0%	99.3%	99.4%	0.854	0.445	1.478				
	10	UFMAj/MFMA	0.653	0.049	-0.065	0.622	0.625	1.040	77.7%	94.0%	94.3%	100.0%	99.5%	99.9%	0.814	0.393	1.087				
		UFMAs/MFMA	0.651	0.049	-0.065	0.628	0.625	1.040	78.5%	94.0%	94.3%	100.0%	99.5%	99.9%	0.818	0.393	1.087				
0.2	5	UFMAj*	0.575	0.014	0.069	0.700	0.667	1.223	84.1%	93.8%	93.3%	100.0%	99.5%	99.3%	0.821	0.446	1.501	Comparison	μ_1	μ_2	μ_3
		UFMAs**	0.592	0.014	0.069	0.709	0.667	1.223	84.9%	93.8%	93.3%	100.0%	99.5%	99.3%	0.854	0.446	1.501	MFMA vs UFMAj	2.1%	2.1%	2.0%
		MFMA***	0.569	0.015	0.069	0.693	0.661	1.211	83.5%	93.1%	92.3%	100.0%	99.4%	99.3%	0.804	0.437	1.471	MFMA vs UFMAs	4.5%	2.1%	2.0%
	10	UFMAj	0.636	0.051	-0.048	0.619	0.625	1.045	78.1%	93.9%	95.2%	100.0%	99.6%	99.9%	0.788	0.394	1.094				
		UFMAs	0.651	0.052	-0.047	0.628	0.625	1.045	78.5%	93.9%	95.2%	100.0%	99.6%	99.9%	0.818	0.394	1.094	MFMA vs UFMAj	3.5%	3.3%	2.7%
		MFMA	0.634	0.053	-0.053	0.608	0.615	1.030	77.1%	91.8%	93.7%	100.0%	99.4%	99.9%	0.772	0.381	1.065	MFMA vs UFMAs	6.2%	3.3%	2.7%
0.5	5	UFMAj	0.489	0.017	0.072	0.690	0.668	1.224	87.9%	94.3%	94.5%	100.0%	99.6%	99.6%	0.715	0.447	1.503	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.592	0.017	0.072	0.709	0.668	1.224	84.9%	94.3%	94.5%	100.0%	99.6%	99.6%	0.854	0.447	1.503	MFMA vs UFMAj	2.3%	2.5%	2.3%
		MFMA	0.482	0.018	0.069	0.682	0.660	1.210	87.2%	92.9%	93.4%	100.0%	99.6%	99.5%	0.697	0.436	1.468	MFMA vs UFMAs	7.5%	2.5%	2.3%
	10	UFMAj	0.544	0.051	-0.020	0.610	0.626	1.046	82.2%	94.2%	94.8%	100.0%	99.6%	100.0%	0.668	0.394	1.094				
		UFMAs	0.651	0.051	-0.020	0.628	0.626	1.046	78.5%	94.2%	94.8%	100.0%	99.6%	100.0%	0.818	0.394	1.094	MFMA vs UFMAj	3.8%	3.9%	3.3%
		MFMA	0.542	0.051	-0.027	0.599	0.613	1.028	81.1%	93.0%	94.0%	100.0%	99.5%	100.0%	0.652	0.379	1.058	MFMA vs UFMAs	9.0%	3.9%	3.3%

*UFMAj: Univariate fixed-effects meta-analysis joint model **UFMAs: Univariate fixed-effects meta-analysis separate model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.7 40% MAR in one outcome (μ1)

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0.8	5	UFMAj**	0.289	0.022	0.067	0.712	0.668	1.225	93.6%	95.3%	95.5%	100.0%	99.7%	99.9%	0.590	0.447	1.504	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.592	0.022	0.067	0.709	0.668	1.225	84.9%	95.3%	95.5%	100.0%	99.7%	99.9%	0.854	0.447	1.504	MFMA vs UFMAj	2.9%	3.9%	3.8%
		MFMA****	0.288	0.029	0.075	0.701	0.655	1.201	92.8%	94.5%	94.8%	100.0%	99.6%	99.9%	0.575	0.430	1.449	MFMA vs UFMAs	2.0%	3.9%	3.7%
	10	UFMAj	0.329	0.043	0.008	0.630	0.626	1.046	90.6%	94.0%	94.6%	100.0%	100.0%	100.0%	0.505	0.394	1.095				
		UFMAs	0.651	0.043	0.008	0.628	0.626	1.046	78.5%	94.0%	94.6%	100.0%	100.0%	100.0%	0.818	0.394	1.095	MFMA vs UFMAj	4.5%	6.0%	5.4%
		MFMA	0.331	0.046	0.006	0.615	0.607	1.017	88.5%	92.4%	93.2%	99.9%	99.6%	99.9%	0.488	0.370	1.035	MFMA vs UFMAs	3.8%	6.0%	5.4%
RA_corr*	5	UFMAj	0.378	0.019	0.067	0.692	0.668	1.224	91.5%	94.1%	94.0%	100.0%	99.7%	99.5%	0.622	0.447	1.502	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.592	0.019	0.067	0.709	0.668	1.224	84.9%	94.1%	94.0%	100.0%	99.7%	99.5%	0.854	0.447	1.502	MFMA vs UFMAj	2.4%	2.7%	2.3%
		MFMA	0.373	0.024	0.070	0.684	0.659	1.209	90.6%	93.6%	94.2%	100.0%	99.6%	99.4%	0.607	0.435	1.467	MFMA vs UFMAs	6.8%	2.7%	2.3%
	10	UFMAj	0.427	0.048	-0.041	0.613	0.626	1.045	86.9%	94.0%	94.4%	100.0%	99.8%	100.0%	0.558	0.394	1.094				
		UFMAs	0.651	0.048	-0.041	0.628	0.626	1.045	78.5%	94.0%	94.4%	100.0%	99.8%	100.0%	0.818	0.394	1.094	MFMA vs UFMAj	3.8%	4.2%	3.3%
		MFMA	0.427	0.049	-0.046	0.601	0.612	1.028	85.8%	93.3%	93.9%	100.0%	99.5%	100.0%	0.544	0.377	1.059	MFMA vs UFMAs	8.3%	4.2%	3.3%

* RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.8 20% MAR in all outcomes

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.301	0.364	0.277	0.724	0.567	1.034	93.6%	88.5%	94.3%	100.0%	99.5%	100.0%	0.615	0.454	1.145				
		UFMAs/MFMA	0.298	0.365	0.273	0.726	0.568	1.036	93.9%	89.1%	94.2%	100.0%	99.5%	100.0%	0.616	0.457	1.149				
	10	UFMAj/MFMA	0.355	0.530	0.285	0.643	0.344	0.884	89.9%	89.2%	93.3%	100.0%	99.9%	100.0%	0.643	0.399	0.863				
		UFMAs/MFMA	0.356	0.344	0.288	0.645	0.532	0.887	89.9%	90.1%	93.3%	100.0%	99.8%	100.0%	0.645	0.401	0.869				
0.2	5	UFMAj**	0.295	0.361	0.425	0.723	0.566	1.039	93.5%	88.2%	93.3%	100.0%	99.6%	100.0%	0.609	0.451	1.260	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.298	0.365	0.434	0.726	0.569	1.044	93.9%	87.7%	93.0%	100.0%	99.6%	100.0%	0.616	0.457	1.279	MFMA vs UFMAj	2.0%	2.1%	2.0%
		MFMA****	0.292	0.362	0.424	0.715	0.560	1.028	92.9%	87.5%	93.3%	100.0%	99.6%	99.9%	0.597	0.444	1.237	MFMA vs UFMAs	3.0%	3.1%	3.0%
	10	UFMAj	0.349	0.352	0.373	0.641	0.529	0.886	90.5%	88.9%	93.1%	100.0%	99.9%	100.0%	0.533	0.404	0.924				
		UFMAs	0.356	0.354	0.387	0.645	0.532	0.891	89.9%	89.5%	92.7%	100.0%	99.9%	100.0%	0.543	0.409	0.943	MFMA vs UFMAj	3.0%	3.4%	2.8%
		MFMA	0.349	0.353	0.371	0.631	0.520	0.873	88.9%	87.7%	92.7%	100.0%	99.8%	100.0%	0.521	0.395	0.901	MFMA vs UFMAs	4.1%	4.5%	3.8%
0.5	5	UFMAj	0.272	0.344	0.389	0.714	0.560	1.028	93.5%	88.8%	92.9%	100.0%	99.8%	100.0%	0.584	0.432	1.208	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.298	0.368	0.433	0.726	0.569	1.045	93.9%	89.3%	91.9%	100.0%	99.5%	100.0%	0.616	0.459	1.279	MFMA vs UFMAj	2.3%	2.4%	2.2%
		MFMA	0.263	0.344	0.381	0.706	0.553	1.016	92.6%	88.5%	92.0%	100.0%	99.8%	100.0%	0.568	0.424	1.178	MFMA vs UFMAs	5.4%	5.6%	5.4%
	10	UFMAj	0.325	0.340	0.350	0.634	0.524	0.876	91.4%	89.3%	92.5%	100.0%	99.9%	100.0%	0.508	0.390	0.890				
		UFMAs	0.356	0.358	0.406	0.645	0.533	0.891	89.9%	89.4%	91.4%	100.0%	99.9%	100.0%	0.543	0.412	0.959	MFMA vs UFMAj	3.3%	3.6%	3.0%
		MFMA	0.315	0.335	0.343	0.624	0.514	0.863	89.9%	89.4%	91.4%	100.0%	100.0%	100.0%	0.488	0.377	0.863	MFMA vs UFMAs	6.4%	6.7%	6.1%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.8 20% MAR in all outcomes

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0.8	5	UFMAj**	0.241	0.300	0.314	0.708	0.555	1.018	94.0%	90.1%	93.7%	100.0%	99.8%	100.0%	0.559	0.398	1.134	Comparison	μ_1	μ_2	μ_3
		UFMAs***	0.298	0.356	0.411	0.726	0.569	1.044	93.9%	88.7%	92.3%	100.0%	99.5%	100.0%	0.616	0.451	1.259	MFMA vs UFMAj	2.7%	2.9%	2.6%
		MFMA****	0.235	0.302	0.313	0.698	0.547	1.004	93.7%	89.8%	93.3%	100.0%	99.8%	100.0%	0.543	0.390	1.106	MFMA vs UFMAs	7.6%	7.7%	7.4%
	10	UFMAj	0.285	0.310	0.293	0.628	0.520	0.840	91.7%	90.9%	92.1%	100.0%	99.9%	100.0%	0.476	0.366	0.868				
		UFMAs	0.356	0.347	0.410	0.645	0.533	0.961	89.9%	89.3%	90.2%	100.0%	99.8%	100.0%	0.543	0.404	0.891	MFMA vs UFMAj	3.8%	4.4%	3.6%
		MFMA	0.277	0.305	0.282	0.616	0.508	0.806	90.6%	89.0%	90.9%	100.0%	99.9%	100.0%	0.457	0.351	0.852	MFMA vs UFMAs	8.6%	9.1%	8.4%
RA_corr*	5	UFMAj	0.259	0.329	0.394	0.710	0.558	1.028	93.3%	90.0%	92.8%	100.0%	99.7%	100.0%	0.571	0.420	1.212	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.298	0.362	0.432	0.726	0.569	1.044	93.9%	89.7%	92.4%	100.0%	99.7%	100.0%	0.616	0.455	1.275	MFMA vs UFMAj	2.4%	2.4%	2.2%
		MFMA	0.252	0.332	0.389	0.701	0.552	1.017	92.8%	89.3%	92.5%	100.0%	99.7%	100.0%	0.556	0.414	1.185	MFMA vs UFMAs	6.7%	6.2%	5.1%
	10	UFMAj	0.306	0.326	0.338	0.630	0.523	0.877	91.2%	91.1%	92.4%	100.0%	99.8%	100.0%	0.491	0.379	0.883				
		UFMAs	0.356	0.351	0.388	0.645	0.533	0.890	89.9%	90.8%	92.2%	100.0%	99.8%	100.0%	0.543	0.407	0.944	MFMA vs UFMAj	3.4%	3.7%	2.9%
		MFMA	0.292	0.321	0.328	0.619	0.513	0.864	90.8%	90.2%	92.3%	100.0%	99.8%	100.0%	0.469	0.366	0.854	MFMA vs UFMAs	7.7%	7.4%	5.8%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

Table E.9 40% MAR in all outcome

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3				
0	5	UFMAj/MFMA	0.595	0.684	0.576	0.703	0.551	1.005	83.9%	74.7%	91.9%	100.0%	98.5%	100.0%	0.848	0.772	1.342				
		UFMAs/MFMA	0.592	0.683	0.574	0.709	0.556	1.013	84.9%	75.6%	91.6%	100.0%	98.5%	100.0%	0.854	0.776	1.357				
	10	UFMAj/MFMA	0.663	0.624	0.648	0.624	0.516	0.858	76.4%	75.9%	86.1%	100.0%	99.5%	100.0%	0.829	0.655	1.156				
		UFMAs/MFMA	0.651	0.620	0.622	0.628	0.518	0.865	78.5%	76.1%	87.6%	100.0%	99.5%	100.0%	0.818	0.653	1.135				
0.2	5	UFMAj*	0.591	0.679	0.691	0.702	0.550	1.012	84.0%	73.4%	89.3%	100.0%	98.4%	100.0%	0.842	0.764	1.501	Comparison	μ_1	μ_2	μ_3
		UFMAs**	0.592	0.681	0.700	0.709	0.556	1.021	84.9%	74.8%	89.7%	100.0%	98.4%	100.0%	0.854	0.773	1.534	MFMA vs UFMAj	2.3%	2.4%	2.2%
		MFMA***	0.588	0.677	0.681	0.694	0.544	1.000	82.8%	72.5%	88.4%	100.0%	98.4%	100.0%	0.827	0.755	1.464	MFMA vs UFMAs	4.2%	4.3%	4.1%
	10	UFMAj	0.659	0.618	0.705	0.622	0.515	0.860	77.5%	77.1%	85.9%	100.0%	99.7%	100.0%	0.822	0.647	1.237				
		UFMAs	0.651	0.617	0.704	0.628	0.519	0.869	78.5%	78.2%	86.0%	100.0%	99.7%	100.0%	0.818	0.650	1.249	MFMA vs UFMAj	5.3%	5.8%	4.8%
		MFMA	0.640	0.621	0.674	0.605	0.499	0.839	73.0%	72.5%	81.8%	100.0%	98.7%	100.0%	0.776	0.635	1.158	MFMA vs UFMAs	6.9%	7.2%	6.7%
0.5	5	UFMAj	0.564	0.647	0.630	0.692	0.543	0.997	84.6%	76.6%	90.3%	100.0%	99.1%	100.0%	0.798	0.713	1.391	Comparison	μ_1	μ_2	μ_3
		UFMAs	0.592	0.671	0.673	0.708	0.556	1.021	84.8%	75.8%	89.6%	99.9%	98.2%	99.9%	0.853	0.760	1.495	MFMA vs UFMAj	2.7%	2.9%	2.6%
		MFMA	0.558	0.644	0.614	0.683	0.535	0.984	84.5%	74.8%	88.3%	100.0%	99.0%	100.0%	0.778	0.702	1.346	MFMA vs UFMAs	6.7%	6.9%	6.6%
	10	UFMAj	0.622	0.600	0.651	0.613	0.507	0.847	78.5%	76.6%	85.4%	100.0%	99.7%	100.0%	0.762	0.617	1.142				
		UFMAs	0.651	0.612	0.698	0.628	0.519	0.868	78.5%	77.5%	85.3%	100.0%	99.6%	100.0%	0.818	0.644	1.242	MFMA vs UFMAj	5.3%	6.1%	4.8%
		MFMA	0.590	0.610	0.609	0.596	0.491	0.827	74.9%	70.6%	81.6%	99.9%	99.0%	99.8%	0.703	0.614	1.055	MFMA vs UFMAs	9.6%	10.3%	9.3%

*UFMAj: Univariate fixed-effects meta-analysis joint model **UFMAs: Univariate fixed-effects meta-analysis separate model ***MFMA: Multivariate fixed-effects meta-analysis

Table E.9 40% MAR in all outcome

Patient level correlation	No. Studies	Method	Bias			SE_{sim}			Coverage			Power			MSE			Borrowing of Strength			
			μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	μ_1	μ_2	μ_3	Comparison	μ_1	μ_2	μ_3
0.8	5	UFMAj**	0.525	0.595	0.577	0.673	0.527	0.967	86.3%	78.6%	90.9%	100.0%	99.6%	100.0%	0.728	0.632	1.267				
		UFMAs***	0.592	0.660	0.668	0.709	0.556	1.019	84.9%	76.7%	89.8%	100.0%	98.4%	100.0%	0.854	0.744	1.485	MFMA vs UFMAj	3.4%	3.6%	3.3%
		MFMA****	0.519	0.596	0.568	0.661	0.518	0.951	84.5%	76.8%	89.2%	100.0%	99.6%	100.0%	0.707	0.624	1.226	MFMA vs UFMAs	12.9%	13.2%	13.0%
	10	UFMAj	0.569	0.580	0.601	0.596	0.492	0.823	80.1%	77.1%	87.4%	100.0%	99.8%	100.0%	0.679	0.579	1.040				
		UFMAs	0.651	0.610	0.725	0.628	0.518	0.868	78.5%	77.2%	85.1%	100.0%	99.5%	100.0%	0.818	0.641	1.280	MFMA vs UFMAj	5.7%	6.4%	5.0%
		MFMA	0.552	0.581	0.575	0.578	0.476	0.803	78.3%	74.6%	83.9%	99.9%	99.4%	100.0%	0.639	0.564	0.975	MFMA vs UFMAs	15.0%	15.5%	14.5%
RA_corr*	5	UFMAj	0.549	0.631	0.633	0.684	0.540	0.999	85.4%	76.0%	89.1%	100.0%	99.4%	100.0%	0.769	0.690	1.398				
		UFMAs	0.592	0.665	0.665	0.709	0.556	1.021	84.9%	75.6%	88.9%	100.0%	98.3%	99.9%	0.854	0.751	1.485	MFMA vs UFMAj	2.9%	3.0%	2.6%
		MFMA	0.541	0.628	0.619	0.674	0.532	0.986	84.7%	75.4%	88.3%	100.0%	98.9%	100.0%	0.747	0.677	1.356	MFMA vs UFMAs	9.7%	8.7%	6.7%
	10	UFMAj	0.596	0.590	0.666	0.606	0.503	0.849	79.3%	76.8%	85.8%	100.0%	99.7%	100.0%	0.722	0.601	1.165				
		UFMAs	0.652	0.611	0.697	0.628	0.518	0.868	78.4%	76.6%	86.2%	100.0%	99.9%	100.0%	0.819	0.643	1.240	MFMA vs UFMAj	5.6%	6.1%	4.8%
		MFMA	0.572	0.595	0.646	0.588	0.488	0.828	76.7%	74.1%	82.3%	99.9%	99.1%	99.7%	0.673	0.592	1.104	MFMA vs UFMAs	12.1%	11.4%	8.9%

*RA_corr: TJC&SJC: 0.67 TJC&Pain: 0.55 SJC&Pain: 0.38 **UFMAj: Univariate fixed-effects meta-analysis joint model ***UFMAs: Univariate fixed-effects meta-analysis separate model ****MFMA: Multivariate fixed-effects meta-analysis

REFERENCES

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't BMJ (1996) 312.7023.71.
2. Davidoff F, Haynes B, Sacket DL, Smith R. Evidence based medicine. A new journal to help doctors identify the information they need. BMJ (1995), Volume 310.
3. Stolberg HO, Norman G, Trop I. Randomized controlled trials. Am J Roentgenol. (2004) 183 (6): 1539–44.
4. Moher D, Schulz K, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet (2001), 357, 1191-4.
5. Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? JAMA (1998). 280, 280-2.
6. Maxwell L, Santesso N, Tugwell PS, Wells GA, Judd M, Buchbinder R. Method guidelines for Cochrane Musculoskeletal Group systematic reviews. The Journal of Rheumatology (2006); 33:11.
7. Ryan R, Hill S, Broclain D, Horey D, Oliver S, Prictor M. Cochrane Consumers and Communication Review Group. Study Design Guide. (June 2013).
8. Helfenstein U. Data and models determine treatment proposals—an illustration from meta-analysis. Postgrad Med J (2002); 78:131-134.
9. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1. www.cochrane-handbook.org, 19 November 2008.
10. Woodward M. Epidemiology Study Design and Data Analysis. Chapter 12 - Meta-Analysis. Third edition (2013). CRC Press.
11. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med (2002), 21, 1539-58.
12. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall/ CRC, (1991).
13. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. 3rd ed. Boca Raton (2004): Chapman & Hall / CRC.
14. Moher D, Liberati A, Tetzlaff J, Altman DG, and the PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Intern Med. (2009); 151(4):264-269.
15. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials, 7 (1986), pp. 177–188.
16. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in Medicine (2012), 31 (20). pp. 2179-2195.
17. Jackson D, Riley RD, White IR. Multivariate meta-analysis: Potential and promise. Statist. Med. (2011), 30 2481–2498.

18. Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomised trials: a review of methods used in practice. Clin Trials (2005), 209-17.
19. Ashley P Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. Clinical Trials (2009); 6: 16–27.
20. Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. Stat Med. (2001); 20:2219–41.
21. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet (2016); 388: 2023–38.
22. David L, Scott DL, Frederick Wolfe F, Huizinga TWJ. Rheumatoid arthritis. Lancet (2010); 376: 1094–1108.
23. Zhang Y, Jordan JM. Epidemiology of Osteoarthritis. Clin Geriatr Med. (2010) August; 26(3): 355–369.
24. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum (1993); 36:729-40.
25. Aletaha D, Neogi T, Silman A, et al. The 2010 American College of Rheumatology/European League Against Rheumatism Classification Criteria for Rheumatoid Arthritis. Arthritis Rheum (2010).
26. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. Trials (2012), 13:132.
27. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, von Elm E, Gamble C, Ghersi D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One (2008), 3:e3081.
28. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ (2010), 340:c365.
29. Clarke M. Standardising outcomes for clinical trials and systematic reviews. Trials (2007), 8:39.
30. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials (2007), 8:38.
31. Schmitt J, Spuls PI, Thomas KS, Simpson E, Furue M, Deckert S, Dohil M, Apfelbacher C, Singh JA, Chalmers J, Williams HC, on behalf of the HOME initiative collaborators. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. J Allergy Clin Immunol. (2014) Oct; 134(4):800-7.
32. Van der Heijde DM, Van't Hof M, Van Riel PL, Van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol (1993); 20:579e81.

33. Dwan K, Gamble C, Williamson PR, Kirkham JJ, the Reporting Bias Group. Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias — An Updated Review. PLoS ONE 8(7), (2013).
34. Song F, Parekh S, Hooper L, Loke YK, Ryder J, et al.. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 14 (2010).
35. Rothstein HR, Sutton AJ, Borenstein M. Publication Bias in Meta-Analysis. Prevention, Assessment and Adjustments: Wiley (2005).
36. Frosi G, Riley RD, Williamson PR, Kirkham JJ. Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews. Journal of Clinical Epidemiology, (2015).
37. Kirkham JJ, Altman DG, Williamson PR. Bias Due to Changes in Specified Outcomes during the Systematic Review Process. PLoS ONE 5(3), (2010).
38. Dwan K, Kirkham JJ, Williamson PR, Gamble C. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. BMJ Open. (2013); 3(6).
39. Kirkham JJ, Boers M, Tugwell P, Clarke M, Williamson PR. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. Trials (2013); 14:324.
40. Dwan K, Gamble C, Ruwanthi KD, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: a tutorial. Trials (2010), 11:52.
41. Katchamart W, Trudeau J, Phumethum V, Bombardier C. Methotrexate monotherapy versus methotrexate combination therapy with non-biologic disease modifying anti-rheumatic drugs for rheumatoid arthritis. Cochrane Database of Systematic Reviews 2010, Issue 4. Art. No.: CD008495.
42. Mottaghi P, Karimzade H. Does chloroquine decrease liver enzyme abnormalities induced by methotrexate in patients with rheumatoid arthritis? J Res Med Science (2005); 10(3): 135–8.
43. Felson T, LaValley MP. The ACR20 and defining a threshold for response in rheumatic diseases: too much of a good thing. Arthritis Research & Therapy, (2014), 16:101.
44. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials. JAMA. (2004); 291(20):2457-2465.
45. Van der Heijde DMFM, Jacobs JWG. The original DAS and the DAS28 are not interchangeable: comment on the articles by Prevoo et al. Arthritis Rheum (1998); 41:942-5.
46. Suarez-Almazor ME, Belseck E, Shea B, Homik J, Wells GA, Tugwell P. Antimalarials for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.

47. Suarez-Almazor ME, Spooner C, Belseck E. Azathioprine for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2009), Issue 1. Published by John Wiley & Sons, Ltd.
48. Suarez-Almazor ME, Spooner C, Belseck E, Shea B. Auranofin versus placebo in rheumatoid arthritis (Review). The Cochrane Collaboration (2009), Issue 1. Published by John Wiley & Sons, Ltd.
49. Suarez-Almazor ME, Belseck E, Shea B, Tugwell P, Wells GA. Cyclophosphamide for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
50. Wells GA, Haguenauer D, Shea B, Suarez-Almazor ME, Welch V, Tugwell P, Peterson J. Cyclosporine for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
51. Clark P, Tugwell P, Bennett KJ, Bombardier C, Shea B, Wells GA, Suarez-Almazor ME. Injectable gold for rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
52. Suarez-Almazor ME, Belseck E, Shea B, Tugwell P, Wells GA. Methotrexate for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
53. Suarez-Almazor ME, Belseck E, Spooner C. Penicillamine for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2011), Issue 10. Published by John Wiley & Sons, Ltd.
54. Suarez-Almazor ME, Belseck E, Shea B, Tugwell P, Wells GA. Sulfasalazine for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
55. Osiri M, Shea B, Welch V, Suarez-Almazor ME, Strand V, Tugwell P, Wells GA. Leflunomide for the treatment of rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
56. Ortiz Z, Shea B, Suarez-Almazor ME, Moher D, Wells GA, Tugwell P. Folic acid and folinic acid for reducing side effects in patients receiving methotrexate for rheumatoid arthritis (Review). The Cochrane Collaboration (2009), Issue 1. Published by John Wiley & Sons, Ltd.
57. Navarro-Sarabia F, Ariza-Ariza R, Hernandez-Cruz B, Villanueva I. Adalimumab for treating rheumatoid arthritis (Review). The Cochrane Collaboration (2009), Issue 9. Published by John Wiley & Sons, Ltd.
58. Maxwell L, Singh JA. Abatacept for rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 1. Published by John Wiley & Sons, Ltd.
59. Mertens M, Singh JA. Anakinra for rheumatoid arthritis (Review). The Cochrane Collaboration (2009), Issue 1. Published by John Wiley & Sons, Ltd.

60. Ruiz Garcia V, Jobanputra P, Burls A, Cabello JB, Gálvez Muñoz JG, Saiz Cuenca ESC, Fry-Smith A. Certolizumab pegol (CDP870) for rheumatoid arthritis in adults (Review). The Cochrane Collaboration (2011), Issue 2. Published by John Wiley & Sons, Ltd.
61. Lethaby A, Lopez-Olivo MA, Maxwell L, Burls A, Tugwell P, Wells GA. Etanercept for the treatment of rheumatoid arthritis (Review). The Cochrane Collaboration (2013), Issue 5. Published by John Wiley & Sons, Ltd.
62. Blumenauer BBTB, Judd M, Wells GA, Burls A, Cranney A, Hochberg MC, Tugwell P, Lopez-Olivo MA. Infliximab for the treatment of rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
63. Singh JA, Noorbaloochi S, Singh G. Golimumab for rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
64. Singh JA, Beg S, Lopez-Olivo MA. Tocilizumab for rheumatoid arthritis (Review). The Cochrane Collaboration (2010), Issue 7. Published by John Wiley & Sons, Ltd.
65. Kirwan JR, Bijlsma JWJ, Boers M, Shea B. Effects of glucocorticoids on radiological progression in rheumatoid arthritis (Review). The Cochrane Collaboration (2009), Issue 1. Published by John Wiley & Sons, Ltd.
66. Ward JR, Williams HJ, Boyce E, Egger MJ, Reading JC, Samuelson CO. Comparison of auranofin, gold sodium thiomalate and placebo in the treatment of rheumatoid arthritis. Am J Med (1983); 133-137.
67. Johnsen V, Borg G, Trang LE, Berg E, Brodin U. Auranofin (SK&F) in early rheumatoid arthritis: Results from a 24- month double blind, placebo-controlled trial. Scand J Rheumatol (1989); 18:251–260.
68. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. BMJ (2010); 340:c332.
69. Saini P, Loke KY, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. BMJ (2014); 349.
70. Williamson PR, Gamble C. Application and investigation of a bound for outcome reporting bias. Trials (2007), 8: 9.
71. Copas JB, Dwan KM, Kirkham JJ, Williamson PR. A model based correction for outcome reporting bias in meta-analysis. Biostat (2014) 15 (2): 370-383.
72. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, Williamson PR. The impact of outcome reporting bias on systematic reviews. British Medical Journal (Research) (2010); 340:c365.
73. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Biostatistics (2008), 9, 1, pp. 172–186.
74. Wei Y, Higgins PT. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. Stat Med. (2013) Mar 30; 32(7): 1191–1205.

75. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Statistics in Medicine 2002; 21:589–624.
76. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ (2010); 340:c221.
77. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, Rovers M. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. PLoS Med 12(7) (2015).
78. White IR. Multivariate random-effects meta-analysis. STATA Journal (2009); 9 (1): 40–56.
79. Riley RD, Price MJ, Jackson D, Wardle W, Gueyffier F, Wang J, Staesseng JA, White IR. Multivariate meta-analysis using individual participant data. Res. Syn. Meth. (2015), 6 157–174.
80. Jackson D, White IR, Price MP, Copas J, Riley RD. Borrowing of strength and study weights in multivariate and network meta-analysis. Stat Methods Med Res. (2015) Nov 6.
81. White IR, Carpenter J, Evans S, Schroter S. Eliciting and using expert opinions about dropout bias in randomized controlled trials. Clinical Trials (2007); 4: 125–139.
82. Rubin DB, Little RJ. Statistical analysis with missing data. Hoboken, NJ: J Wiley & Sons; (2002).
83. Powney M, Williamson PR, Kirkham JJ, Kolamunnage-Dona R. A review of the handling of missing longitudinal outcome data in clinical trials. Trials (2014), 15:237.
84. Carpenter JR, Kenward M. Missing data in clinical trials - a practical guide. 2008, Birmingham, UK: National Institute for Health Research, Publication RM03/JH17/MK.
85. SAS/STAT(R) 9.2 User's Guide, Second Edition.
86. Laird NM; Ware JH. Random-Effects Models for Longitudinal Data. Biometrics (1982) 38, 963-974 963–974.
87. Schafer J.L. Analysis of Incomplete Multivariate Data. New York: Chapman and Hall (1997).
88. Copas J, Jackson D. A bound for publication bias based on the fraction of unpublished studies. Biometrics (2004); 60(1):146-53.
89. Egger M, Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. British Medical Journal; (1997) 315: 629–634.
90. Jackson D, White I, Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. Stat Med (2012); 31:38.
91. Williamson PR, Altman DG, Blazeby JM, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. J Health Serv Res Policy (2012); 17.
92. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PLoS One (2013); 8(7): e66844.